Performance Effects of Computer-Based Multitasking Behavior

by

RACHEL F. ADLER

A dissertation submitted to the Graduate Faculty in Computer Science in partial fulfillment of the requirements for the degree of Doctor of Philosophy, The City University of New York

2012

This manuscript has been read and accepted for the
Graduate Faculty in Computer Science in satisfaction of the
dissertation requirement for the degree of Doctor of Philosophy.

PROFESSOR RAQUEL BENBUNAN-FICH

_______________ ____

Date Chair of Examining Committee

PROFESSOR THEODORE BROWN

_______________ ____

Date Executive Officer

PROFESSOR MATT HUENERFAUTH
PROFESSOR RICHARD HOLOWCZAK
PROFESSOR MARK SILVER

____

Supervisory Committee

THE CITY UNIVERSITY OF NEW YORK

Abstract

PERFORMANCE EFFECTS OF COMPUTER-BASED MULTITASKING BEHAVIOR

by

RACHEL F. ADLER

Adviser: Professor Raquel Benbunan-Fich

This research examines multitasking from the perspective of human-computer interaction (HCI). Multitasking is defined as the performance of multiple tasks concurrently. In a computer-based environment, users generally switch between multiple computer-based tasks either due to a personal decision to break from the current task or due to an external interruption, such as an electronic notification. This dissertation describes an in-depth empirical study, using a laboratory setting with different numeric, verbal, and visual computer-based tasks. Six hundred and thirty six subjects were randomly assigned into three conditions: *discretionary multitasking*, where participants were allowed to decide when and how often to switch tasks, *forced multitasking*, where participants were forced to switch tasks at certain allotted times, and *non-multitasking*, where participants performed the tasks sequentially and were not allowed to multitask. In order to investigate performance effectiveness (accuracy) and performance efficiency (productivity), participants' overall accuracy and productivity scores were compared across conditions. The results suggest that during difficult tasks, subjects who were forced to multitask had the lowest accuracy. In addition, those subjects in the forced multitasking

condition who felt the primary task was difficult had lower accuracy than those who felt the task was easy. This was not true in the other two conditions. Receiving interruptions during a difficult task impacted not only their primary task, but their secondary tasks as well. In the discretionary multitasking condition, the more subjects decided to multitask, the lower their accuracy scores. In fact, an additional analysis revealed that high multitaskers not only performed worse than low and medium multitaskers in the discretionary condition, but actually had the worst performance than subjects in any other condition. Medium multitaskers, however, had the highest productivity scores. While multitasking in that case was considered the best in terms of efficiency, it was not true in terms of effectiveness. Therefore, discretionary multitasking gives the illusion of high performance. Furthermore, this study also explored why people chose to multitask and the impact that had on performance. The results of this study can assist HCI researchers in developing a more comprehensive understanding of user multitasking which can lead to better interface designs.

Acknowledgements

I would like to express my appreciation to the Doctoral Student Research Grant that funded my pilot experiment. I was able to pay 100 subjects for participating in my research. This was useful in forming the experimental design for the actual laboratory experiment. Over 300 subjects in the actual experiment received monetary compensation from PSC-CUNY Research Grant # 62552-00 40 given to Professor Raquel Benbunan-Fich. I am very grateful for the above two grants which allowed me to include many subjects in my research. I would also like to thank the many subjects who took time to participate in the experiment. Without them this project could not have happened.

I would like to thank several people without whom I could not have completed my dissertation. The first and foremost is my advisor, Professor Raquel Benbunan-Fich. I am not able to thank her enough. Her constant guidance over email, the phone, and in person kept me on track throughout this process. I really appreciate the time she spent answering my numerous questions. I know I could not have had a better advisor and completing my dissertation would not have been possible without her. I owe her more than I can express.

I would like to thank my committee members, Professor Matt Huenerfauth, Professor Richard Holowczak, and Professor Mark Silver, who frequently provided me with wonderful feedback and were always very supportive.

I would also like to thank my family. My mother-in-law and father-in-law (who unfortunately is no longer around to see the finished product) always had faith and thought more of me than I did. To my many sisters, brothers-in-law, and sister-in-law, for providing me with a place to escape and comic relief throughout this process. To my parents, who encouraged me to pursue my Ph.D. and constantly guided me with helpful suggestions and ideas. And finally, my husband, who I definitely could not have gone through this process without: for constantly proof-reading my work, testing my experiment, and listening to me brainstorm. I know I can never repay you for your help and love.

Lastly, my dissertation is dedicated to my daughter Emily, the bright spot of every day. Although sometimes she can be more of a distraction to my work, this multitasking makes my life worth living and provides me with never-ending happiness, love, and laughter. I dedicate my dissertation to Emily with the message that anything is possible; I hope that you will be able to accomplish all of your dreams.

Table of Contents

List of Tables

List of Figures

CHAPTER 1. INTRODUCTION

Human-Computer Interaction (HCI) deals with the interaction between a user and a computer through an interface. Studying HCI is important in order to design computer applications that are simple for people to use. Hochheiser and Lazar (2007) reviewed the HCI literature pertaining to societal issues. They identified two main influences which stimulate HCI responses: the needs of various users (such as individuals, families, communities, governments, and businesses), and technological advances. Using these two parameters as a guideline, this research discusses the user and technology, by examining how users are in fact using their systems. A designer needs to understand how users intend to use their systems in order to best design applications for them.

Typically, computers are used by people in the workplace and at home. Users can be on the computer for reasons as simple as searching for information online, answering email, or playing games. There are two broad purposes associated with the use of computers: *hedonic*, i.e. leisure-oriented reasons such as playing games, and *utilitarian*, which are productivity-oriented purposes like doing homework. People often combine multiple unrelated tasks in a single computer session.

Since multitasking is very prevalent in today's world, understanding why people multitask and how it affects performance can enable HCI researchers to design effective systems. This dissertation presents a research study which examines the effects computer-based multitasking has on user performance. It compares the performance of subjects who can decide when and how often to multitask with the performance of subjects who are forced to immediately multitask and subjects who are not multitasking. For those subjects who were allowed to decide whether or not to multitask, an in-depth analysis was performed analyzing how often they switched, why they switched, and whether those decisions impacted their performance.

The first part of this dissertation is the background. First, a discussion of human-computer interaction is provided followed by a discussion of the different ways users perform multiple computer-based tasks. Then, relying on concepts discussed by Norman (1988) and a review of the recent literature on HCI, the dissertation adopts a computer science perspective to understanding human multitasking. The dissertation then presents research questions related to computer-based multitasking and describes the hypotheses for this research study.

The second part of the dissertation provides the methodology and results. The methodology chapter consists first of a description of a pilot experiment that was performed and then goes into the details of the actual experiment. Next, the results of the data analyses of the empirical study are described. The dissertation then presents a discussion of the results, in addition to the implications this has for practice, research, and interface designs, followed by the conclusion.

CHAPTER 2. BACKGROUND

2.1 Human-Computer Interaction

Human-Computer Interaction (HCI) is defined as "a discipline concerned with the design, evaluation and implementation of interactive computing systems for human use and with the study of major phenomena surrounding them" (ACM SIGCHI 1992). The ACM SIGCHI Curricula for Human-Computer Interaction (1992) also states that from a computer science viewpoint HCI focuses on the interaction between one or more humans and one or more machines, specifically highlighting application design and the engineering of human interfaces.

HCI is defined by the confluence of computer science and other disciplines such as psychology, artificial intelligence, linguistics, anthropology, and sociology. This is because a computer can also be thought of as a tool that impacts social interaction or as a personal assistant (Draper and Norman 1986). While computer scientists who study HCI focus on how to design applications, psychologists have played an important role in HCI research, since their primary concern is the users of the computer systems themselves and understanding the needs of the user is a significant part of their research (Card et al. 1983). HCI can also be articulated in terms of the technology (such as the hardware and software used), the social system (such as people and psychology), and the interaction between the two (Zhang et al. 2004).

Douglas et al. (2002) advocate including human-computer interaction into the computer science curriculum and argue that 50% of code is programmed for user interfaces. John and Morris (1993: 49) argue that "many computer systems do not realize their full potential to enable users to accomplish tasks easily and efficiently." Furthermore, they state that computer science originally emphasized the importance of techniques that would make computer systems faster and cheaper, such as having more efficient algorithms. However, with computers already running quickly and at lower prices, computer science needs to focus more on helping users complete tasks easily. People are using computers to accomplish tasks, so in addition to technology, it is important for computer scientists to understand the task that a user wants to perform. Figure 2-1 shows that understanding human capabilities and tasks are just as important as understanding computer technologies when designing computer systems.

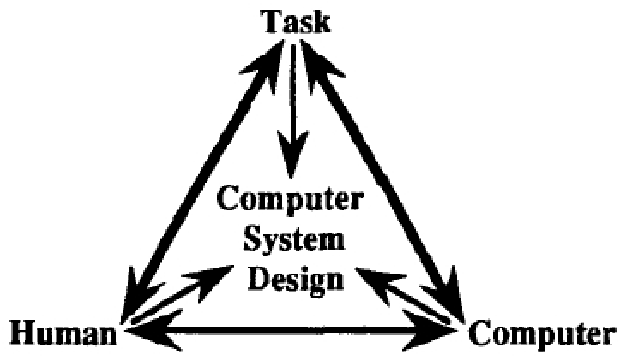


Figure 2-1. Task-Human-Computer Triad (John and Morris 1993)

The next section discusses user interfaces from the point of view of HCI and how user interfaces can be an effective medium to support humans performing tasks.

2.1.1 User Interfaces

In *The Design of Everyday Things*, Norman (1988) discusses how users are not at fault when using the latest technology unsuccessfully, but rather the designer is. Bannon (1986) uses the term "idiot-proof systems" to describe easy-to-use interfaces. The term implies that a system must be designed in a way to protect them from humans' mistakes.

Norman (1988) examines the sources of users difficulties by analyzing the structure of an action. The cycle of an action is defined in terms of both execution and evaluation. Execution is defined as what gets done while evaluation is comparing what happened with what we wanted to have happen. Figure 2-2 shows Norman's *Seven Stages of Action*. For the *Stages of Execution*, it begins on top with the goal, i.e. the desired state. The goal is the intention to do a specific action. This intention is then translated into an action sequence which will be performed in order for the intention to be fulfilled. Nothing actually happens until it is executed. The *Stages of Evaluation* begins with a perception of the world. The perception gets interpreted according to one's expectations and then evaluated based on the user's intentions and goals. Norman's framework identifies two problems for users: the gulf of execution and the gulf of evaluation. The gulf of execution is concerned with the performance of actions, i.e. how well a user is able to carry out the intended actions. A user may intend to do one thing, but the actual sequence of actions might be more complicated than his intentions. The gulf of evaluation refers to the effort that is spent assessing the state of the system, i.e. the outcome of the actions, for example if a system looks like it is working but in fact it is not.

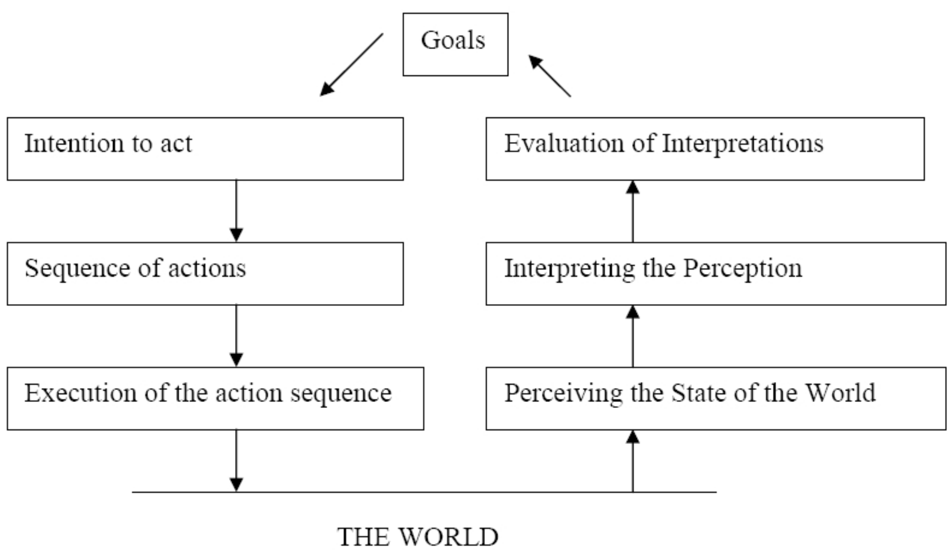


Figure 2-2. Norman's Seven Stages of Action (Norman 1988)

This paper will focus on how the Stages of Execution (Figure 2-3) change when the user decides to perform multiple computer-based tasks.

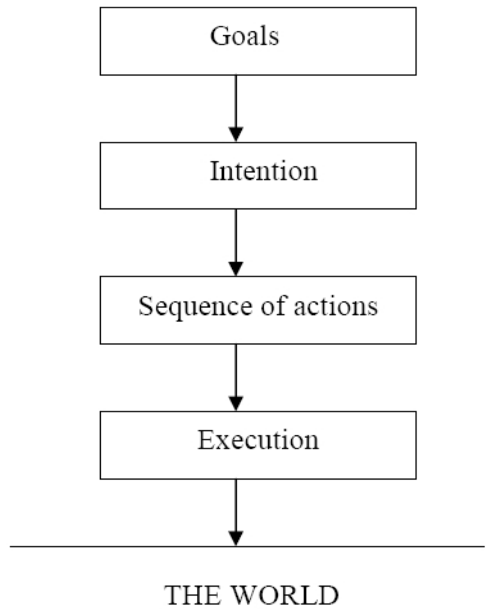


Figure 2-3. Norman's Stages of Execution (Norman 1988)

There are several factors for computer scientists to keep in mind when designing user interfaces. It is crucial to know your user by talking to them and watching them. The interface needs to be designed such that it has maximum usability, which can be thought of as how easily a system can be used. By contrast, functionality consists of having functions in the system that match what the users want the system to do (Goodwin 1987). Shneiderman (1992) argues that designers, managers, and programmers need to fight for the user by making sure commands and terminology on interfaces are consistent, displays are in readable formats, and errors messages are understandable. The designer also needs to allow ease of use while still maintaining high performance (Dix et al. 2004). Having ease of use with powerful software can be a trade-off. The conflict arises when trying to make powerful systems that are also easy to use.

In HCI, knowing who will be interacting with the application is an important part of designing an effective application. Designers need to understand that users today are more likely to want systems that allow them to easily multitask and deal with interruptions in the form of instant messenger or email notifications. The purpose of examining this area is to demonstrate how variations in users' intentions, tasks, and applications produce different interaction settings. It is crucial for a designer to understand potential patterns of interaction to avoid poor results.

2.2 Performing Multiple Tasks

Since personal technology is advancing, HCI is no longer only concerned with one person using one interface, but how users typically use their computers, which involves switching back and forth between multiple tasks (Johnson et al. 2003). This area will be reviewed by emphasizing the users' execution of actions, i.e. how the user switches between multiple tasks and the implications for research.

In a typical computer environment, a user is interacting with a computer for a specific purpose. Early personal computer environments only allowed users to work on one task at a time. However, with the advent of modern windows-based operating systems, users can work on more than one task concurrently. Individuals' approach to time usage can be *monochronic* or *polychronic*. Monochronic people perform tasks one at a time, while polychronic types of people will perform several tasks at once (Bluedorn et al. 1992; Zhang et al. 2005). Polychronic individuals are more likely to engage in multitasking than monochronic individuals. Lee (1999) argues that just as individuals can arrange their time when dealing with events and tasks in a monochronic or polychronic way, tasks and events can also be monochronic or polychronic. Events that occur irregularly without a fixed schedule are polychronic, while those that arrive in a regular sequence are monochronic. Polychronic individuals are likely to combine actions of different tasks and perform them polychronically, while monochronic individuals will tend to carry multiple tasks in sequence.

Salvucci et al. (2009) have coined the term *multitasking continuum* which refers to the amount of time spent on one task before switching to another task. They classify task switching in two categories: concurrent or sequential. Examples with human-based multitasking are shown in Figure 2-4. On the left-hand side are tasks that are rapidly switched between and therefore occur pretty much at the same time, such as listening and note-taking or driving and talking. On the right-hand side of the continuum are tasks they view as closer to sequential task switching. These tasks have longer intervals of time before switching occurs, such as cooking and reading a book.

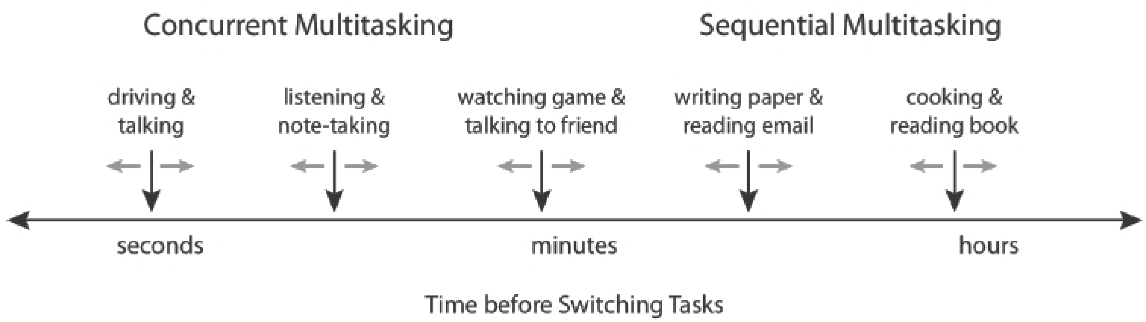


Figure 2-4. The Multitasking Continuum (Salvucci et al. 2009)

This paper classifies performing multiple tasks with three different strategies: parallel, sequential, and interleaving (Bluedorn et al. 1992). When a user is multitasking on a computer it is very hard to find cases of pure parallel behavior. One example of pure parallel behavior on the computer can be someone who is surfing the web with music playing in the background, using a software application such as iTunes or Media Player. Figure 2-5 presents an illustration of parallel task performance.

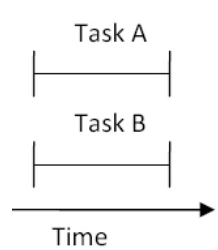


Figure 2-5. Parallel Strategy for Performing Multiple Tasks

Sequential switching occurs when tasks are performed one after the other; in such a way that one task is started after the previous one has been completed. An example of sequential behavior would be a student who is working on their homework on their computer and then finishes and goes to check out a social networking site. Figure 2-6 shows sequential task performance.

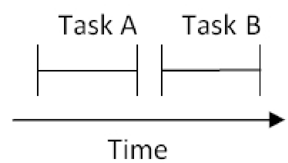


Figure 2-6. Sequential Strategy for Performing Multiple Tasks

Interleaving is when a task is intentionally stopped or interrupted in order to perform a different task (Payne et al. 2007). This third approach, interleaving, is the way in which most people carry out multiple tasks on their computers. When interleaving, a user can stop an ongoing task and switch to a secondary task, and then resume the first one. Figure 2-7 shows a case of interleaving with two tasks, where the primary task (A) is stopped to complete another (B). Upon the completion of B, task A is resumed.

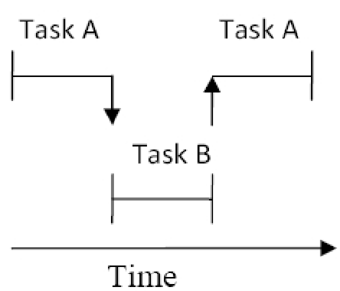


Figure 2-7. Interleaving Strategy for Performing Multiple Tasks

For simplicity, the previous figures show only two tasks. However, the strategies are applicable to contexts with multiple tasks. Interleaving, in particular, can occur in many different ways, depending on how many tasks are performed and the patterns of interruptions and returns to previous tasks.

Norman's Stages of Execution can be adapted to the case of multiple tasks. In the original model, there is a goal, which is the state the user wants to achieve, and the intention to do some action becomes an action sequence to complete that goal. In the case of multiple unrelated tasks, there are multiple goals, which represent the tasks, and in order to complete these goals/tasks there are multiple action sequences. Figure 2-8 shows a modification in Norman's Stages of Execution with multiple sequences of actions in order to complete many tasks instead of one task.

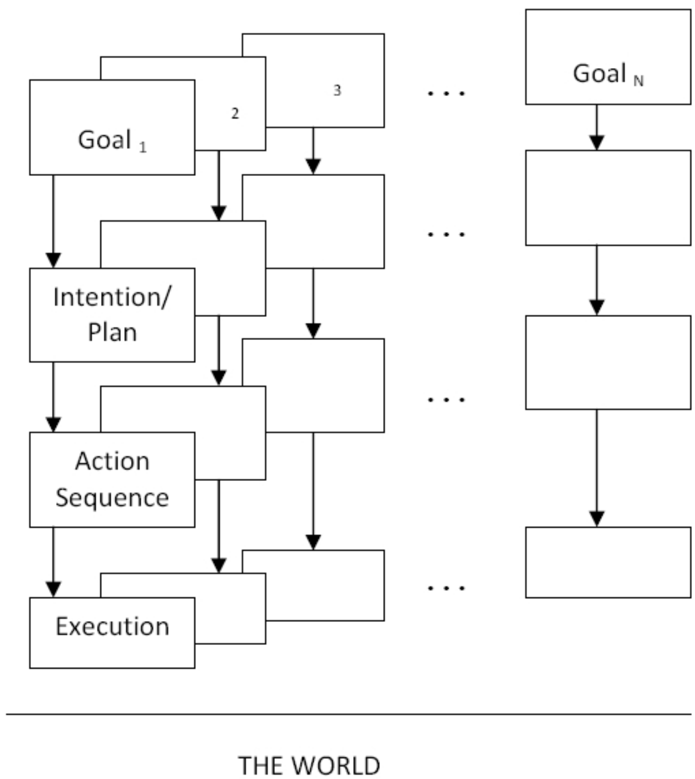


Figure 2-8. Norman's Stages of Execution: Sequences of Actions Leading to the Completion of Multiple Tasks.

If the tasks are completed in a parallel situation, where the execution of the action sequences occurs at the same time, then Figure 2-8 can be viewed from top to bottom. In this way, all the actions are done at the exact same time. If the tasks are performed sequentially, Figure 2-8 should be viewed from left to right. First there is one goal/intention/action sequence/execution, and then another. As time increases, each task is being executed one after the other.

When interleaving occurs, there are switches occurring before the completion of tasks. From the action level, this means that these action sequences can occur in many ways. For example, Figure 2-9 shows a view from the action level showing the user interleaving the tasks. In this case, a user interrupts Action Sequence$_1$ switches to Action Sequence$_2$, then switches to Action Sequence$_3$ and then resumes Action Sequence$_1$ (i.e. the primary task). This interleaving of tasks, i.e. multitasking, is how users general switch between tasks.

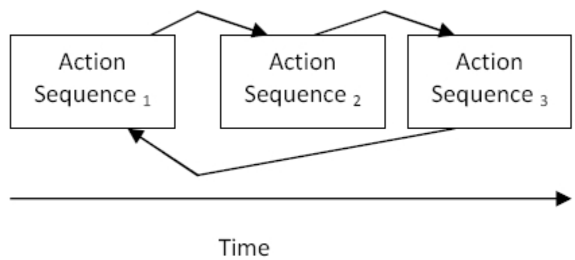


Figure 2-9. Interleaving Action Sequences

A summary of the studies reviewed in this section is provided in Table 2-1.

Table 2-1. Summary of Research for Performing Multiple Tasks

Author(s), Year	Research Question/Objective	Method	Results
Bluedorn et al. (1992)	Explains the difference between monochronic and polychronic individuals.	Conceptual	Understanding the difference between monochronic and polychronic people can help one better manage themselves and their organizations.
Johnson et al. (2003)	How can we reduce our limited understanding of collaborative and multiple tasks?	Introduction to a special issue on multiple and collaborative tasks.	Borrow methods from other disciplines.
Lee (1999)	To study how information technology affects temporalities of work.	E: brought a new system, KTNET, into two work environments and analyzed how that changed temporalities of work.	Information technology creates a temporal symmetry between work groups that are interacting.
Salvucci et al. (2009)	Unified theory of multitasking.	Conceptual	Created a new theory, building on three previous theories, that spans continuous and sequential forms of multitasking.
Zhang et al. (2005)	What are the differing control strategies and performance of monochronic vs. polychronic individuals?	E: 42 university students were tested whether they were monochronic or polychronic and then they performed two tasks.	Monochronic individuals performed tasks one after the other and polychronic at the same time. The polychronic individuals had fewer errors.

*E=Empirical

2.3 Definition of Multitasking

"Multitasking" is an ambiguous term since it can refer to many different scenarios. Different types of multitasking can be classified as follows:

1. *CPU-based multitasking* occurs when the computer is multitasking, i.e. the CPU is switching from one program to another so quickly that it appears as if programs are being executed at the same time.
2. *Human-based multitasking* occurs when a human is performing multiple tasks at the same time. This can refer to either:
    a. *General multitasking* which refers to any situation when a person is performing many manual, motor, or perceptual tasks at once, such as talking on the phone, eating lunch, and checking email.
    b. *Computer-based multitasking* which occurs when a computer user switches windows back and forth solely on the computer in his/her performance of multiple tasks. This can refer to either one of the following:
        i. *Application-based multitasking* which occurs when people switch between applications on their computer.
        ii. *Task-based multitasking* which refers to people switching between unrelated tasks which are performed entirely on the computer.

Multitasking in this paper refers to human-based multitasking with particular emphasis on computer-based tasks. While this research focuses on computer-based multitasking, other multitasking studies look at other platforms as well. For example, in a recent study, Ophir et al. (2009) examined media multitasking, i.e. users who switched between different media such as print, television, computer-based video, music, and the telephone.

In computer-based multitasking, it is important to distinguish between task-based multitasking and application-based multitasking. One could have a single task that uses multiple programs or many tasks which all use the same program (Cypher 1986). When a person is using the computer, s/he can use many different applications either to complete different aspects of the same task, or to carry out multiple tasks. In the latter case, a user is often working on one (primary) task and then switching to work on a different (secondary) task, and then eventually resuming the primary task.

Application-based multitasking can involve two types of switching: a switch to a different application or a switch within the same application. Switching to a different application occurs when a user decides to change to a completely new application (A_D). Alternatively, a user can switch to another instance of the same application, i.e. a window switch. Switches that occur between windows (or tabs), of the same application are labeled A_S. For example, switching to another tab on a Web browser or another word document in Microsoft Word would be a case of A_S.

Like application-based multitasking, task-based multitasking can be classified into switching to a different task (T_D) or within the same task (T_S), in other words, *between-task switching* and *within-task switching* (Wild et al. 2004). *Between-task switching* occurs when a user switches windows on his computer in order to change to a completely different task. *Within-task switching* occurs when a user switches windows on his computer for the purpose of completing the same task. For example, a student using Google to look up information for an essay assignment might be using two applications: a Web browser and Microsoft Word. This is an example of a within-task switch since even though a different application is being used, the primary goal remains the same (i.e. writing the essay).

The combination of task switches with potential applications changes produces four possible situations regarding switches (See Figure 2-10). For example, when the underlying task is the same the user can switch to either another window of the same application (T_SA_S) or to an entirely different application (T_SA_D). On the right hand side of the table, the user is switching to a different task and either the user is switching to another window within the same application (top, T_DA_S) or to a different application (bottom, T_DA_D).

TASKS

APPLICATIONS		SAME	DIFFERENT
	SAME	T_SA_S	T_DA_S
	DIFFERENT	T_SA_D	T_DA_D

Figure 2-10. Four Switching Cases

Wild et al. (2004) mention that there needs to be a good way to model multitasking, since task modeling models one person and one task, and while those studying events and interruptions understand that people are polychronic and work on more than one task at a time, they do not have a good representation of task. They also discuss how between-task relationships need to be understood better. Benbunan-Fich et al. (2009b; 2011) have defined lean and rich metrics for multitasking. The leanest measure indicates whether or not one is multitasking, while the richest measure proposed is the percentage to which one is multitasking, i.e. the percentage of times one switched to an application not related to one's current task. This is calculated by the number of switches between different tasks divided by all switches.

Understanding how people multitask by changing applications and/or tasks and handle interruptions such as instant messenger (IM) and email notifications is important for HCI researchers, since they are concerned with designing applications that will improve users' efficiency and effectiveness. Sometimes users switch tasks only after a task is done, in a sequential fashion, while others voluntarily switch back and forth to check their email or respond to an instant message. The latter is becoming a more common method of computer usage. Determining how people use their computers is beneficial for producing standards which will assist in making computer applications simpler and more effective.

CHAPTER 3. REVIEW OF MULTITASKING LITERATURE

Not only do computer scientists need to understand users' tasks, but they also need to consider how users switch between them in order to accomplish multiple tasks. In *User Centered Design System: New Perspectives on Human-Computer Interaction,* Norman and Draper (1986) include a section on user activities, i.e. users interleaving their activities, since as they say, "to our knowledge, these aspects of real tasks have seldom been discussed, let alone supported … by computer systems" (Page 240). McCrickard et al. (2003c: 511) state "Although the general study of supporting multiple activities with interface design is not a new area of concern within HCI, it has not been an area of intense and cohesive focus."

To this end, we have examined six of the most influential HCI journals[1] (*International Journal of Human Computer Interaction*, *Human Computer Interaction*, *Interacting with Computers*, *Behaviour and Information Technology, International Journal of Human-Computer Studies,* and *ACM Transactions on Computer-Human Interaction)*, using no timeframe[2], and searching for keywords such as "multitasking", "multitask", "multi-tasking", "task switch", or "task switching" and have found 14 articles, including two introductions to special issues (Appendix A lists the articles from HCI journals). In addition, we have examined all of the proceedings for the Conference on Human Factors in Computing Systems (CHI) looking for the same keywords as before and have found 22 articles that discussed computer-based multitasking[3]

---

[1] These journals were selected from a list of HCI periodicals provided by the Human-Computer Interaction Resource Network (http://www.hcirn.com/).

[2] We performed this search in May of 2009 and used this date as the end time.

[3] We focused our search on articles related to computer-based multitasking, i.e. where a person is on a personal computer and multitasking. Therefore, some large multitasking topics, such as driving and multitasking (for example talking on the phone while driving) and multitasking on mobile devices were not included in this review of computer-based multitasking. In addition, in order to narrow the scope of the search, keywords such as

(Appendix B lists the articles from CHI proceedings). The relatively low number of multitasking articles suggests that human multitasking is an emerging area in HCI and that there is still potential to contribute to the HCI-multitasking literature.

HCI researchers are uniquely positioned to explore human multitasking due to their quest to improve human-computer interaction and create interfaces to help people successfully carry out their tasks and by extension, multitask.

This section reviews different areas that are addressed in the multitasking literature, not only mentioning those HCI articles and proceedings mentioned above but also incorporating some research done outside of the HCI literature in order to get a more comprehensive understanding of users' multitasking patterns.

3.1 Multitasking Setting

With personal computers getting faster, people are using more applications at the same time to perform unrelated activities. It is very common to have Microsoft Word or Excel and a web browser open simultaneously, in order to deal with email while working on something else. In addition, with computer screens becoming larger, the number of open windows increases and users are engaged in even more multitasking (Czerwinski et al. 2006).

Multitasking can occur in different locations, such as at home, school, or in the workplace. For example, Bell et al. (2005) discuss multitasking in the workplace, such as people checking their e-mail during a meeting or using IM during group work. Due to its portability,

---

"interruptions" or "notifications" were not included and therefore there could be other articles focused on these topics which are not counted in this number.

computers are also becoming integrated into many meetings and classrooms and the next few sections will discuss the impact multitasking has on these types of sessions.

3.1.1 Multitasking in the Workplace

Bell et al. (2005) mention that multitasking can help work groups since members can work on more than one task at the same time. Nevertheless, it can hinder performance as well since it may be disruptive since people multitasking can become distracted. In addition, they argue that perceptions of others multitasking, whether positive or negative, will effect a member's satisfaction and therefore the effectiveness of the group. Kleinman (2009) studied people's perception of multitasking and found that more people felt that meetings were less productive due to multitasking.

One of the applications that people constantly switch to is email. Renaud et al. (2006) argue that users do not realize the disruptive power of email and Bellotti et al. (2005) indicate that many people experience email overload. Renaud et al. (2006) analyzed the log results of 6 users for about 320 session hours, and found that 23.7% of the time, about 76 hours, was spent doing email. In the majority of email sessions, the users were just checking their email without responding. They argue that constantly checking email reduced the amount of time that users spent on work related applications and therefore lowers the user's rate of productivity. They also mentioned that although people thought that they checked their email about every hour, actual usage tracking showed that it was actually closer to every five minutes.

Su and Mark (2008) found that while email has more back and forth messages, the total amount of time spent on it was shorter than synchronous communication, and IM messages were even shorter. Therefore, they argue that asynchronous communication can be the best option when multitasking and under time pressure.

Wasson (2004) examined multitasking in virtual teams and found that multitasking can be beneficial for those settings, since employees can work on other things while not detracting their attention away from the meeting. Czerwinski et al. (2004) studied how workers in an office dealt with task switches over the course of a week. They found that on average people switched tasks 50 times per week and found it difficult to return to older tasks.

In another study on multitasking in the workplace, González and Mark (2004) found that people spend three minutes on a task before they switch to another task. They define a working sphere as a set of events which share a common goal. After removing non-significant interruptions (such as signing something), they found that individuals spent approximately 12 minutes in a working sphere before moving to another. Mark et al. (2005) found that in the workplace people spend very little time on one task before switching, and 57% of their work is interrupted. Isaacs et al. (2002) looked at IM usage in the workplace. They found frequent users of IM had more multitasking, i.e. they switched out of the message window more often.

3.1.2 Multitasking in the Classroom

It is becoming quite common to find students using their laptops in the classroom. Benbunan-Fich and Truman (2009) investigated how students use their laptops during lectures. They found that students who multitasked with a laptop during a lecture spent 76% of their time on distracting activities.

Hembrooke and Gay (2003) discuss multitasking in the classroom as well. During one lecture, one group of students was allowed to use laptops in the classroom, and engage in web browsing and instant messaging, and was tested on the lecture immediately afterwards. The second group was not allowed laptops during the lecture, and was then tested immediately after as well. The results indicated that students who used a laptop had more difficulty remembering lecture content than those who did not.

3.1.3 Multitasking at Home

Nowadays, most people have PCs in their home and multitask between many different activities on their computer. Crook and Barrowcliff (2001) examined students' multitasking habits in their dorms and found that students usually had several applications open at the same time and switched between them. In addition, almost every participant had a sound or video player running in the background.

In another study, Weisz et al. (2007) combined watching online videos with instant messaging in order to make it a more social activity. They found that concurrent chatting had a positive influence on social relationships with the others in the group and people chatted despite being distracted from the video. Groups with chat enjoyed poor cartoons more than groups without chat.

While people can be on the computer for hedonic reasons, such as watching a movie or instant messaging a friend, or utilitarian purposes, such as doing homework, in a study on home computer usage, Benbunan-Fich et al. (2009a) found that students usually combined these two purposes in a single session.

Multitasking is a common occurrence, whether one is at home, school, or work. People can multitask for different reasons such as being bored of a task or receiving an instant message from a friend. The next section therefore discusses the different drivers which cause one to multitask.

An overview of the studies reviewed in section 3.1 is provided in Table 3-1.

Table 3-1. Summary of Research for Multitasking Setting

Author(s),Year	Research Question/Objective	Method	Results
Bell et al. (2005)	Understanding the social implications of multitasking.	Conceptual: Focusing on task switching, such as checking e-mail during a meeting or using IM during group work.	Model of social perceptions of multitasking.
Bellotti et al. (2005)	Examine the phenomenon of email overload.	E: Email filters on seven participants for several weeks.	To embed task management directly in email.
Benbunan-Fich et al. (2009a)	Analyzing home computer usage among college students.	E: 133 self-reported student logs of a computer session.	Most users performed both hedonic and utilitarian computer tasks during the same session.
Benbunan-Fich and Truman (2009)	Understanding how people use their laptops during meetings.	E: Monitoring logs of 67 students using laptops in over 28 80-min lecture sessions.	76% of task switching was on distracting activities while 13% was on compliant activities.
Crook and Barrowcliff (2001)	What are the patterns of home computer usage among college students?	E: Examined random sample of students using computers in study bedrooms.	Activities were usually multitasked. Computers were used more for entertainment than study.
Czerwinski et al. (2004)	Examined people interleaving multiple tasks when interrupted.	E: Diary study of 11 Microsoft Windows users.	Participants performed many task switches and encountered a lot of interruptions. On average people switched tasks 50 times per week and found it difficult to return to older tasks.
Czerwinski et al. (2006)	Evaluate usability issues for large displays.	E: Used a tool to examine window management activity in order to discover patterns of activity for different sized displays, such as the number of open windows.	One issue that comes up is task management problems. As screen sizes become larger, the number of open windows increases and users engage in more multitasking. Therefore, better task management mechanisms will become essential. Also, window management problems occur, since larger displays can cause notifications to pop up in unexpected places.
Gonzalez and Mark (2004)	To understand how information workers manage multiple activities.	E: Field study, observing and interviewing information workers. 14 people over seven months. Each was observed for 3.5 days.	People spend about three minutes on a task before switching. They use working spheres to explain how people organize their work. People spent about 12 minutes in a working sphere before switching.
Hembrooke and Gay (2003)	The effects of multitasking in the classroom.	E: 44 students. During one lecture, one group of students used laptops and one did not. They were both tested on the lecture immediately afterwards.	Students who used a laptop had more difficulty remembering lecture content than those who did not.

*E=Empirical

Table 3-1 Continued. Summary of Research for Multitasking Setting

Author(s),Year	Research Question/Objective	Method	Results
Isaacs et al. (2002)	How people use IM.	E: Collected 23 weeks of IM conversations in a work context.	Frequent users of IM had more multitasking (left the message window more often). IM was used only occasionally to set up interactions in other media. This suggests that IM integration with voice or video may not be that important.
Kleinman (2009)	How are face-to-face meetings impacted by technological multitasking?	E: Subjects from fieldwork at a software development company and survey from a sample of 40 information workers.	There was a perceived loss of productivity when using laptops during meetings. The type of meeting was the strongest determinant for when multitasking occurs.
Mark et al. (2005)	Examine fragmented work.	E: 24 information workers. Data: Observation, long interviews, and shadowing	People spend very little time on one task before switching, and 57% of their work is interrupted.
Renaud et al. (2006)	Explores the effects of email.	E: Monitored 6 people using email and other activities in an academic setting for 3 months.	Constant monitoring of e-mail reduces productivity.
Su and Mark (2008)	Focusing on communication chains to better understand multitasking.	E: Observations by shadowing 19 people in large corporation for over 18 months.	Multitasking involves not only switching among tasks but also switching communication partners and media. Email had more back and forth communication, but was of shorter duration than synchronous communication, and IM messages were even shorter. Therefore, when people are under time pressure, asynchronous communication can be the best option. When external interruptions occurred they had more back and forth communication, and more media switches and more organizational switches. More organizational switching (i.e. conversations outside of one's department) was associated with higher levels of stress.
Wasson (2004)	Examine multitasking during virtual teams.	E: Field study - observed and videotaped 12 employees (4 teams) for about 8 months.	Multitasking could enhance employee productivity when properly managed.
Weisz et al. (2007)	Examine chatting while watching a video.	E: 85 participants into 30 groups. In a lab setting where they showed cartoons on separate computers. Some had a chat feature and some did not. Some were groups of friends.	Chat had a positive influence on social relationships with the others in the group and people chatted despite being distracted. Groups with chat enjoyed poor cartoons more than groups without chat.

*E=Empirical

3.2 Drivers of Multitasking

There are two different drivers of multitasking: *external* and *internal* interruptions. An *external interruption* occurs when an event in the environment causes one to switch tasks, while an *internal interruption* comes from one's self, i.e. self-initiated switching (Miyata and Norman 1986). These drivers correspond to two types of human information processing: *Task-driven processing* and *Interrupt-driven processing*. *Task-driven processing* is where one is mostly focusing on one task and ignoring other events, such as reading a book. *Interrupt-driven processing* is where one is constantly switching activities, such as in the workplace. Since a task-driven individual's attention is on a specific task, there is a limit to how much information they can process in a secondary task. These individuals can only do a minute amount of processing of external activities. However when people are easily distracted, i.e. interrupt-driven processing of information, they are sidetracked by external interruptions as well as internal ones (Miyata and Norman 1986).

Depending on the task, an individual's attention will vary. While some tasks may require one to pay close attention, other tasks can cause individuals to be distracted more easily. Therefore, the importance of the task for the user is likely to determine whether the user will follow an interrupt-driven or a task-driven approach.

Although different tasks may cause multitasking behaviors to vary, the next section will discuss what drives people to multitask to begin with. Multitasking occurs when internal interruptions take place or when external ones emerge. The next two sections will explain the difference between these different types of interruptions which cause people to multitask.

3.2.1 Internal Interruptions

Jin and Dabbish (2009) found that users interrupted themselves three times per hour.

They identified seven categories for internal interruptions which explain why a user would switch to another task:

1. *Adjustment* - improving an aspect of the environment to help the primary task.
2. *Break* - frustrated or tired of the primary task.
3. *Routine* - performing a different task due to habit.
4. *Wait* - filling time until s/he can continue with the primary task.
5. *Inquiry* - needs to get more information for the current task.
6. *Trigger* - performing a related task that the primary task reminded the user of.
7. *Recollection* - remembering about another task that had to be done.

In the psychology literature, Payne et al. (2007) experimented with participants performing two similar computer-based tasks and letting subjects switch between the tasks at will. The results of this experiment indicated that people made frequent switches either because tasks were no longer rewarding or because they finished a subgoal and decided to take a break from the current task by attending to another.

Why people multitask is an important question. People may multitask because they are bored or they could multitask if a notification has alerted them to an outstanding task that needs to be addressed. Interruptions need to be examined since interruptions many times cause users to multitask and may have negative effects on performance (Oulasvirta and Saariluoma 2004; Oulasvirta and Saariluoma 2006).

3.2.2 External Interruptions

Examining the HCI journals has shown that the most prevalent topic in multitasking deals with external interruptions and/or notification systems which force a user to switch to another task (McCrickard et al. 2003a; McCrickard et al. 2003b; McCrickard et al. 2003c; McFarlane

2002; McFarlane and Latorella 2002; Oulasvirta and Saariluoma 2004; Oulasvirta and Saariluoma 2006; Trafton et al. 2003). Designing these notification systems to cause interruptions in an effective way is a significant concern since HCI is concerned with designing interfaces efficiently.

Interruptions can occur frequently when using a computer. Email notifications and instant messages are two examples of interruptions. Bailey and Konstan (2006) study interruptions. Their results indicate that when interrupted, users needed more time to finish their primary task. In addition, they made more errors in both tasks and had more annoyance and anxiety than those who were not interrupted.

Iqbal and Horvitz (2007b) found that per hour, a user was interrupted by an average of 4.28 email alerts and 3.21 IM alerts. In addition, a participant spent about 10 minutes on switches caused by alerts, and another 10 to 15 minutes before resuming the task, since they would visit other applications besides the notifying one. They found that 27% of suspended tasks were not resumed for over two hours.

Cutrell et al. (2000) studied instant messages and found that if an interruption had nothing to do with the task it took longer to process the message and to resume the task than when receiving relevant messages. Mark et al. (2008), however, found no difference in the context of interruptions. They did find that people completed interrupted tasks in less time and with no difference in quality, but had to work faster and therefore had more stress, frustration, time pressure, and effort.

In a study of undergraduate students, Speier et al. (2003) found that interruptions helped performance for a simple task, but hurt performance on more complicated tasks. O'Conaill and Frohlich (1995) found that in the workplace, although 64% of the time people gained from an interruption, 41% of the time they never resumed their primary task.

Trafton et al. (2003) examined people's responses to being interrupted. They use the term *interruption lag* to refer to the time the alert arrives until the actual interruption, for example the ringing of the phone. During this time people can quickly prepare for resuming the interrupted task later, such as wrapping up part of their task. They performed an experiment with a primary computer-based task and a secondary one. Participants were randomly assigned into one of two conditions. In the Warning condition, participants received an eight second interruption lag, while in the Immediate condition participants were taken to the second task with no warning. The results indicated that those in the Warning condition had more preparation time and were therefore able to resume the interrupted task faster. Those in the Immediate group, however, became faster than they were previously with practice.

Iqbal and Horvitz (2007a) show the effect conversation has on the multitasking behavior of computer users. They focus on conversation since getting into a conversation with a person can cause more of a disruption than an alert on a computer. This is due to a lack of time to prepare for a task switch when approached by a person wishing to hold a conversation.

Another study considered interruptions on mobile platforms as opposed to on desktops. Although people believe that phones may be more disruptive, their results indicated that instant messages disrupted the web task more. Furthermore, anticipated interruptions allowed better web performance, specifically on a mobile platform. In addition, the task took longer to complete when interrupted on a mobile device than a desktop (Nagata 2003).

This section described what causes multitasking, whether it gets triggered by one's self or externally. Understanding the interruptions that cause multitasking can help when designing interfaces to better support these multitasking users.

A summary of the articles discussed here which pertain to interruptions is provided in Table 3-2.

Table 3-2. Summary of Research for Multitasking Drivers

Author(s),Year	Research Question/Objective	Method	Results
Bailey and Konstan (2006)	Measuring the effects of interruption.	E: 50 subjects randomly assigned to an experimental group, where the primary tasks were interrupted by other tasks, or to a control group, where the other tasks were between the primary tasks (not during).	When interrupted users needed from 3-27% more time to complete the task and performed twice as many errors in the tasks. They also experienced from 31-106% more annoyance and twice as much anxiety.
Cutrell et al. (2000)	Studied interrupting users with instant messages during difficult parts of a task.	E: 9 participants performing a task and interrupted by instant messages.	Interruptions that did not pertain to the task were more disruptive than those that did. They resulted in longer times to process the message and to resume the task.
Iqbal and Horvitz (2007a)	The effects conversation has on the multitasking behavior of computer users.	E: Field study – used a tool which logged user activities and was also able to detect when conversation occurred for 16 people for 2 weeks.	When interrupted by conversation, users usually suspended their computing tasks, and may perform other tasks such as checking email and web searches at a reduced rate. The longer the duration of the task before the interruption and if the window was visible led to faster resumption of the task.
Iqbal and Horvitz (2007b)	Examine the effects of interruption on task switching.	E: Logging tool to capture data on 27 users for over 2 weeks.	Per hour, a user was interrupted by an average of 4.28 email alerts and 3.21 IM alerts. On average, participants spent about 10 minutes on switches caused by alerts, and another 10 to 15 minutes before resuming the task, since they would visit other applications in addition to the notifying one. 27% of suspended tasks resulted in over two hours until they were continued.
Jin and Dabbish (2009)	Examine internal interruptions.	E: Shadowed 13 participants for a total of 11 hours as they completed their work tasks and interviewed them afterwards.	39 self-interruptions on the computer. Participants interrupted themselves 3 times per hour. They identified 7 categories of self-interruption: *adjustment, break, routine, wait, inquiry, trigger,* and *recollection.*

*E=Empirical

Table 3-2 Continued. Summary of Research for Multitasking Drivers

Author(s),Year	Research Question/Objective	Method	Results
Mark et al. (2008)	Does the context of interruptions make a difference?	E: 48 students. The experiment simulating an office environment and participants answered email messages regarding human resource questions. They were interrupted by their “manager” with some questions pertaining to the task and some not pertaining.	There was no difference in context of interruptions. People completed interrupted tasks in less time with no difference in quality, but had to work faster and therefore had more stress, frustration, time pressure, and effort.
Nagata (2003)	Examines interruptions on mobile platforms as opposed to on desktops and whether the interruption was originated by instant messenger or the phone.	E: 9 participants either assigned to mobile or desktop group.	Instant messages disrupt a web task more than the phone. Interruptions that are anticipated allow better web performance, specifically on a mobile platform. In addition, the task took longer to complete when interrupted on a mobile device than a desktop.
O’Conail and Frohlich (1995)	How disruptive are interruptions in the workplace?	E: Observational study - 2 subjects shadowed with a video camera for a week.	64% of the time people gained from an interruption, but 41% of the time they never resumed their primary task.
Oulasvirta and Saariluoma (2004)	Whether interruptions affect memory accuracy.	E: undergraduate students had a 30-s interruption to the main task.	Interruptions had a significant negative impact on memory accuracy.
Oulasvirta and Saariluoma (2006)	How to overcome interruptions without adverse effects.	E: 108 undergraduate and masters students. They had a main task which was interrupted by a 30-s interactive task.	The first three experiments showed that memory of the task was not affected by the interruption. In the fourth, there was less time for the main task, and the interruption had a bigger impact.
Payne et al. (2007)	Understanding how people decide which task to work on when.	E: Different experiments. Testing how participants switched between different tasks.	Participants performed many switches between the two tasks because either tasks were no longer rewarding or because they finished a subgoal.
Speier et al. (2003)	Explores the effects of interruptions on different types of tasks.	E: 136 undergraduate students performed tasks.	Interruptions helped performance for a simple task, but hurt performance on more complicated tasks.
Trafton et al. (2003)	Examine the effects of having preparation time prior to an interruption.	E: Some participants received an 8-s interruption lag, while some did not.	Those in the Warning were able to resume the interrupted task faster. The others were faster with practice; people adapt to disruptions.

*E=Empirical

3.3 Interfaces

Since people are multitasking, interfaces can be designed to assist in that purpose. As discussed, people may multitask because of an internal interruption, such as wanting a break from the current task. In addition, a user may multitask due to an external interruption, such as notification systems, which cause them to switch. Therefore, the next two sections discuss interfaces designed for effective multitasking and interfaces for managing notifications.

3.3.1 Interfaces for Managing Multitasking

There are different ways computer systems can be designed in order to enable users to more smoothly multitask. Reichman (1986) argues that interfaces should mimic topic switches in human conversation in order to better support multitasking. For instance, in conversation people switch topics, but use phrases such as "well anyway" in order to switch back. So too, interfaces need window managers to have similar techniques in order to improve multitasking.

Scupelli et al. (2005) created a different version of IM in order to better facilitate people working on many projects. IM has the advantage that people can multitask, however as it stands it may not help with people switching between many projects and teams efficiently. Therefore, they designed an IM plug-in which shows the active projects and team members for each project.

Matthews et al. (2006) compared two different techniques which could improve multitasking: Semantic content extraction and change detection. Semantic content extraction displayed only the relevant information in a window, such as displaying a progress bar with text to indicate an upload in progress. Change detection indicated when a change took place, such as highlighting a border of a window red to indicate that a change in that window has occurred, such as receiving new email. The results indicated that semantic content extraction improved multitasking performance.

Another good system for multitasking is activity-based computing (ABC) where the computer is organized in terms of activities. For instance, Card and Henderson (1987) mention that task switching can be time-consuming since the user needs to remember the names and locations of files and programs needed for the second task. They therefore provide an environment with different screen-sized workspaces each related to a different task. Similarly, Voida and Mynatt (2009) studied using an activity-based computing system as opposed to the application- and document-centric one of most desktop computing platforms. In this system, for each activity, a user has a new empty virtual desktop. All files that are saved to the desktop become associated with the current activity. It is very likely that this kind of environment will lead to simpler multitasking.

Bardram et al. (2006) also examined ABC. Figure 3-1 shows their environment, where an activity bar replaces the Windows XP Taskbar, since activities are the main focus and not applications. The list of activities is displayed on the toolbar, and the windows used for the resumed activity “ABC CHI Paper” are displayed on the desktop. The results of their study indicated that people found ABC easy to use and likely to be useful.

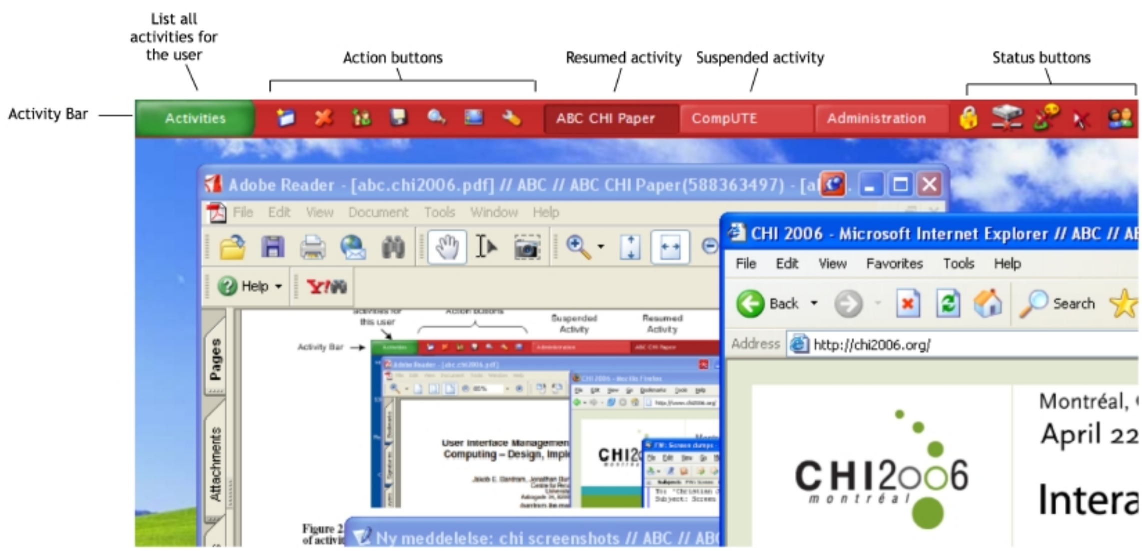


Figure 3-1. Activity-Based Computing Environment (Bardram et al. 2006)

Besides these computer interfaces which can help with multitasking, a big driver of multitasking is notification systems. These can be designed in a way which can cause the interruption to be less disruptive.

3.3.2 Interfaces for Managing Interruptions

How to handle interruptions is a significant Human-Computer Interaction (HCI) issue, since humans are more error-prone when they are interrupted. As a result, multitasking technologies need to be flexible and accommodate different ways of handling interruptions. Furthermore, systems that support effective multitasking have different user interface requirements than single-task systems (McFarlane and Latorella 2002).

When performing a task, there are moments that are better for interruptions (Adamczyk and Bailey 2004; Bailey and Iqbal 2008). Iqbal and Bailey (2006) found that characteristics of the structure of a task can reasonably predict the cost of interruption. This can enable systems to make decisions on when it is better to interrupt. Iqbal and Bailey (2007) examined how to detect breakpoints during tasks. Iqbal and Bailey (2008) scheduled notifications during breakpoints, rather than delivering them immediately, and found that this reduced frustration and reaction time.

McCrickard et al. (2003b) present a model to help in evaluating notifications systems. People want notification systems, such as Instant Messengers or news tickers, to provide information effectively without any undesired interruptions to their primary task. Some notifications systems provide information without disturbing the user by being displayed as sidebars or corner applications. They provide a model for evaluating these different systems based on three factors: interruption, whether the user's attention is shifted to another task, reaction, how fast the user can respond to the notification, and comprehension, whether or not the information is understood quickly or retained. Using IRC, McCrickard et al. (2003a) tested different methods for a notification such as having it fade, ticker, or blast as well as modifying the speed and size. Overall, the best secondary display was one with slow fade.

Karam and Schraefel (2005) studied reducing the effects of interruption. In their experiment they used gestures to support secondary tasks. While participants were completing a primary task they were interrupted and forced to control a secondary task, listening to music. In some cases they used gestures to support this task, such as clockwise circular motion representing "play" and an open handed halt gesture for "stop", and in others they used the keyboard. The task recovery time (the time between the completion of the secondary task and resuming the primary task) was significantly shorter when using gestures.

McFarlane (2002) tested four types of interruptions: immediate, negotiated, mediated, and scheduled. Immediate is where the user gets an interruption for a different task and has to respond right away. Negotiated is when the user can decide whether or not to suspend the task immediately or at some later point. Mediated is when a middle agent decides when to interrupt the task, and scheduled is when the interruptions are scheduled such as every fifteen minutes. The result of the experiment was that in general negotiated multitasking was the best method.

Interruptions cause people to multitask. Sometimes a notification which occurs may be important and therefore multitasking may be beneficial, but at other times the notification may be more of a distraction. Designing notification systems so that they can be helpful without being a hindrance to one's work is an important concern in HCI and in designing effective interfaces. In addition, multitasking interfaces such as the activity-centric ones mentioned, where users can organize their files and desktops in terms of activities can be a very beneficial way to have a friendly atmosphere for multitasking.

3.3.3 Interface Discussion

There are currently some gaps in the HCI literature that should be addressed. Since HCI deals with the interactions between the user and the computer, it requires an understanding of both aspects and an awareness of how people actually use their personal computers nowadays. Because of this dual understanding, HCI researchers are uniquely positioned to investigate human multitasking. Despite the connection between HCI and multitasking, there is currently a paucity of multitasking studies in HCI. Multitasking research, like HCI in general, also deals with the user and computer interacting effectively. However, instead of considering the accomplishment of one task, multitasking research focuses on the user performing multiple tasks, which is more consistent with current user-interaction environments. In addition, since HCI is interested in creating easy-to-use interfaces, it should be more concerned with addressing research on interfaces that support multitasking. For example, notification systems need to be designed in such a way that interruptions do not cause major disruptions.

As the previous sections show, the HCI literature on interfaces can be classified into two categories depending on whether the main objective is to support multiple activities or handle interruptions. Activity-based computing (ABC) appears to be a very good way for individuals to achieve effective multitasking. Having different environments for different tasks can save time when individuals switch between different projects. ABC may help individuals when being forced to multitask due to an interruption. Card and Henderson (1987) point out that a big problem with task switching is that when users are interrupted they might forget what they were previously doing. Therefore, in their system the workspace that was started is left exactly as it was before the interruption. Bardram et al. (2006) found that users felt that ABC would help them handle multiple tasks as well as interruptions. However, while some studies showed that people perceived ABC to be useful, they do not attempt to measure how people actually performed in these environments.

The literature on interfaces and interruptions is mostly focused on how such interruptions can be handled and cause less disruption in the user's ongoing task. For example, McFarlane (2002) found negotiated interfaces to be the best method. By allowing the user to decide when to respond, s/he is free to interleave the new activity associated with the interruption into her/his flow of work. While the literature examines performance with these different interrupting interfaces, the setting is limited to the user working on one task and receiving interruptions. In a typical environment, however, a user might be working on multiple activities.

The goal for studying multitasking in human-computer interaction is to be able to create an interface design which will be an effective system for users. An interface which supports a combination of external and self-initiated interruptions would produce an environment where users are free to work on multiple tasks of their choosing and receive and handle interruptions with minimal disruptions. In our modern window-based computer systems people organize their files using folders. In addition, the taskbar displays the programs that are currently being used. When people are interrupted more files and programs get opened, leaving the desktop unorganized and confusing when a user wants to resume the original task after an interruption has occurred. While ABC uses different workspace environments for each activity, nevertheless it does not take into account the best methods through which people should receive the interruptions themselves.

Very little work in HCI discusses interfaces that deal with both external and internal interruptions. Before we can determine accurately the most useful design, more research needs to be done to jointly examine the two drivers of multitasking: internal and external interruptions and their effects on performance. In this way HCI researchers can determine the most useful design which will encourage people to act as close as possible to optimal performance conditions. While McFarlane (2002) and Bailey and Konstan (2006) examined interruptions, they did not look at self-initiated switching. Other work (Jin and Dabbish 2009; Payne et al. 2007) examined self-initiated switching without exploring external interruptions or investigating performance.

In addition to the lack of multitasking articles in the HCI literature, of those articles most discuss external interruptions and notification systems, and very few address self-initiated interruptions which occur just as often as external interruptions (Gonzalez and Mark 2004). HCI research should include more studies on self-initiated multitasking, since researchers need to understand their users and have systems evolve to easily include simple methods for users to efficiently work on multiple tasks simultaneously.

This research will describe a computer-based experiment comparing the performance of people voluntarily multitasking, forced to immediately multitask, and people who are not multitasking at all. By controlling the tasks and the types of interruptions, researchers can study their effects on performance. For example, if subjects in the forced condition perform the worst, then perhaps interfaces can be designed that have defaults that will minimize interruptions, except when specifically desired by the user.

Performance will be measured using a variety of factors, such as time and accuracy (Bailey and Konstan 2006). Interrupting complex tasks versus simpler tasks will be compared (Speier et al. 2003). Evaluating interruptions in both of those settings can allow us to determine the effect of interruptions in different cases. In addition, examining a discretionary situation where users can decide whether or not to switch and how often can lead to a better knowledge of the drivers of internal interruptions.

Exploring these different situations can help in gaining a better understanding of how and why people multitask along with the effects it has on performance. This research will help in planning a design for interfaces that are easy-to-use and will accurately support actual computer users.

A review of the studies summarized in this section is shown in Table 3-3.

Table 3-3. Summary of Research for Interfaces

Author(s),Year	Research Question/Objective	Method	Results
Adamczyk and Bailey (2004)	Measure the effects of interrupting users at different times.	E: 16 undergraduate and graduate students performed tasks and were interrupted by other tasks.	Designing an attention manager system that could identify moments which were better for interruptions would decrease the disruptive effects of interruptions.
Bailey and Iqbal (2008)	Examine how a user's mental workload changes during task execution.	E: 24 users performed tasks while their pupil dilation was measured.	Some results are that between subtasks, there was a decrease in workload and there was more of a decrease when there was a larger completion of a task.
Bardram et al. (2006)	Using activity-based computing (ABC).	E: Activity bar replaced Window XP taskbar. 16 graduate students tested it and were given a questionnaire and video-taped interview afterwards.	Users found it easy to use and likely to be useful.
Card and Henderson (1987)	Design an interface to help users task switch.	Conceptual	Provide a number of screen-sized workspaces called Rooms. Each room is related to a different task.
Iqbal and Bailey (2006)	Examine how well characteristics of task structure predict Cost Of Interruption (COI), as measured by resumption lag.	E: Experiment 1: costs of interruption were measured by having users interrupted, and calculating the resumption lag. Experiment 2: Explored how well their model predicted COI classes.	The resumption lag values were classified into their COI classes using the cluster information from experiment 1. The majority were predicted correctly. This showed that the three characteristics of task structure (Level, Carryover, and Difficulty of Next Subtask) were able to predict COI with fairly high accuracy. This can allow systems to make more effective decisions about when to interrupt.

*E=Empirical

Table 3-3 Continued. Summary of Research for Interfaces

Author(s),Year	Research Question/Objective	Method	Results
Iqbal and Bailey (2007)	Detecting breakpoints in tasks.	E: 24 observers were asked to review six users' screen interaction videos, identify breakpoints, and enter their reason. Breakpoints with high agreement were considered true.	Models were able to detect breakpoints with reasonably high accuracy.
Iqbal and Bailey (2008)	Present a Notification Management System.	E: six users – using data collection software with two studies, the first to detect breakpoints, the second to schedule notifications then.	The first study showed the system detecting breakpoints, but not their type. The second study showed that rather than delivering notifications immediately, scheduling them at breakpoints reduced frustration and reaction time.
Karam and Schraefel (2005)	Using gestures to support secondary tasks.	E: Participants were given a primary task to complete while their secondary task forced them to control music. In some cases they used gestures to support this task, such as clockwise circular motion represented the play button, and in others they used the keyboard.	The task recovery time (the time between the completion of the secondary task and resuming the primary task) was significantly shorter when using gestures.
Matthews et al. (2006)	Techniques to improve multitasking. They compare semantic content extraction (displays only the most relevant content) and change detection (signals when a change has occurred in a window).	E: 26 users with computer experience. Performed two higher tasks which required waiting for updates and lower priority tasks.	Semantic content extraction improves multitasking performance.
McCrickard et al. (2003a)	Evaluate Notification Systems.	E: Test Interruption-reaction-comprehension (IRC) Experiment1: 70 undergraduates – used ticker (continues horizontally), fade (gradually fades), and blast (places item in display). Experiment2: checks fading and tickering with different speeds and sizes.	Experiment 1: Tickering was the best for comprehension, fading for reaction. Experiment 2: Smaller displays were better for reaction, but more interruptive. Slower displays had more comprehension. Overall, slow fade was the best secondary display.
McCrickard et al. (2003b)	How to evaluate notification systems.	E: Case study presenting a model based on interruption, reaction, and comprehension (IRC) using questionnaires to evaluate Scope (a notification system).	The IRC-based questionnaire supported the actual redesign strategy decisions better than the original questionnaire.

*E=Empirical

Table 3-3 Continued. Summary of Research for Interfaces

Author(s),Year	Research Question/Objective	Method	Results
McFarlane (2002)	How to handle interruptions.	E: 36 participants to compare four methods of interruption.	Negotiation was the best solution.
McFarlane and Latorella (2002)	Why interruption is an important HCI problem, and will continue to grow in importance.	Conceptual	Reviews interruption-management problems and existing design guidance for successful cases and provides a theoretical foundation for improved design guidance for interruption management.
Scupelli et al. (2005)	Using IM to juggle many projects and teams.	E: Experiment comparing regular IM to their version of IM (PVIM) on 20 groups of four people. PVIM shows the active projects and team members for each project.	PVIM was not any more difficult and reduced their workload stress.
Voida and Mynatt (2009)	Study activity-based computing as opposed to the application- and document-centric of most desktop computing interfaces.	E: recruited five mac users to use their activity-based interface.	Users had a positive experience with it.

*E=Empirical

CHAPTER 4. RESEARCH QUESTIONS

The primary focus of the current research was to investigate whether multitasking affects users' performance. This was done by comparing the results of subjects' performance in three conditions:

1. Discretionary Multitasking (DMT) – Users voluntarily switching between tasks.
2. Forced Multitasking (FMT) – Users being interrupted by tasks.
3. Non-Multitasking (NMT) – Users performing tasks sequentially, one following the other.

The goal of this experiment was to try and answer the following questions:

1. Are there differences in performance (speed and accuracy) in the primary and interrupting tasks within these three different conditions?
2. Are the effects of interruptions different for tasks that subjects find easier versus tasks that they consider harder?
3. How often do people switch when able to multitask at their discretion and how does that affect their performance?

The first question attempts to determine whether the condition will affect the subjects' performance. Performance is defined both in terms of speed and accuracy. Speed is measured based on the time spent on a task, while accuracy measures the actual correctness of each task. The second question attempts to answer whether the level of difficulty of a task plays a significant role in its performance when receiving interruptions. The third question considers subjects in the discretionary condition and determines whether there are differences in performance depending on whether they are heavy or lighter multitaskers.

CHAPTER 5. HYPOTHESIS DEVELOPMENT

When users undertake multiple tasks in one computer session, they may decide on their own to switch between unrelated tasks or they might be forced to do so when they are interrupted. A user can be working on a primary task and one or more secondary tasks can force disruptions if they cause the user's attention to shift away from that primary task. McFarlane (2002) compared different methods of interruption and their effects on performance and found that the immediate condition caused people to make the most errors. The *forced* multitasking condition that will be discussed here is similar to McFarlane's (2002) immediate condition. When a user has to deal with mandatory interruptions while performing another task, it can be associated with more errors. Therefore, we propose

H1. Mandatory interruptions will have a negative effect on accuracy.

There is a cognitive cost associated with switching between tasks (Payne et al. 2007). These can account for some of the negative effects of interruptions. There are times that interruptions cause more of a disturbance than other times. Payne et al. (2007) found that one reason a user may choose to voluntarily switch to another task would be after the completion of a subgoal. However, when a user is forced to multitask due to a notification alert, this can cause more of a disturbance if it does not occur after the conclusion of a subtask, but during it. When people can multitask at their discretion this allows them to decide how to perform multiple tasks. This is similar to McFarlane's (2002) negotiated condition where users decided when to perform each task in order to complete them in the allotted timeframe. By removing the constraint that requires the forced handling of a task, users will experience fewer negative effects on performance.

H1a. Those who multitask at their discretion will have greater accuracy than those forced to multitask.

Since, as discussed above, there is a cognitive cost associated with task switching (Payne et al. 2007), those who are not forced to multitask but can perform their tasks sequentially, one following the other, will not have those added costs and therefore will perform better in completing their tasks. Bailey and Konstan (2006) compared those interrupted during a task with those receiving the interruptions between tasks. Their results indicate that when interrupted, users needed more time to finish their primary task, made more errors in both tasks, and had more annoyance and anxiety. In a situation where a user has to complete multiple tasks, performing those tasks one following the other without multitasking can lead to better results than being forced to answer an interruption instantly.

H1b. Those who do not multitask will have greater accuracy than those forced to multitask.

Furthermore, it is reported that when interrupted, people found it very hard to return to previous tasks (Czerwinski et al. 2004). Jin and Dabbish (2009) discuss how when voluntarily switching to another task, users have to spend time recalling the primary task. Becoming interrupted either by oneself or by an external interruption can lead to more time spent on resuming one's train of thought for the primary task. Tasks can be completed much faster when the user does not have to spend additional time recalling what one had been in the middle of working on before they switched tasks.

H2. Those who do not multitask will complete their primary task faster than those who multitask.

H2a. Those who do not multitask will complete their primary task faster than those who are forced to multitask.

H2b. Those who do not multitask will complete their primary task faster than those who multitask at their discretion.

In a discretionary condition, users can switch between different tasks based on their own decisions. Self-initiated switching may have a negative effect on performance (Jin and Dabbish 2009). As discussed, although people tend to perform multiple tasks, simultaneously, it is not without cost (Payne et al. 2007). When a user chooses to switch very often, they are forcing themselves to lose their train of thought and to get distracted. Therefore, constant switching between tasks can increase the error rate.

H3. When multitasking at their discretion, those who decide to switch tasks more often will have lower accuracy than those who switch less often.

Payne et al. (2007) found that people tended to switch to another task when they felt a task was no longer rewarding. Many times people take a break from a task. Deciding to take a break or being forced to take a break can affect a user's performance. Depending on the level of difficulty of a task, people may benefit from being forced to take breaks. While studies have compared performance for those users interrupted versus those who did not receive interruptions, few studies have examined differences which occur based on task difficulty. Speier et al. (2003) found by examining the speed and accuracy of simple and complex tasks that interruptions during a task helped performance with simple tasks, but hurt for more complicated ones. When involved in a complicated task having an interruption can make people lose their concentration and therefore have negative effects. However, if it is an easy task, less cognitive effort is required and therefore switching to a secondary task may not make much of a disturbance. In fact, multitasking in those cases may be beneficial if one can accomplish many goals in the same amount of time. Bedny and Karkowsky (2007) have said that "the concept of complexity is critically important for usability evaluation." They discuss how complex tasks can require more cognitive energy than easier tasks and task performance will be impacted. Therefore, if users are forced to multitask during a complex task, their performance will be negatively impacted. Therefore, we will attempt to verify that

H4. Those forced to multitask during a task they consider harder will perform worse than if they considered the task easier.

Payne et al. (2007) state two reasons why a user might voluntarily switch between tasks: either the user completed a subgoal, or the user decided to switch to a more rewarding task. For example, when performing a harder task, a user may switch to an easier task, since it can be more rewarding. Jin and Dabbish (2009) also examine different situations in which one may decide to multitask, such as when the user needs a break, or the task is taking too long to perform. In an experimental laboratory environment where people have a specific number of tasks that they can switch between, Payne et al.'s (2007) reasons apply. Jin and Dabbish (2009) shadowed individuals as they performed their typical tasks and therefore were able to see many additional reasons which will not apply in an experimental condition, such as a task reminding a user about a different task that needed to be performed. Yet, from both of these studies we see how people switch tasks when they are frustrated with a task and need a break (Jin and Dabbish 2009) or want a more rewarding task (Payne et al. 2007). If a task is hard for a user, they may decide to switch to an easier one, either because they want a break, or they feel that their time is better spent on a task that is more doable. Therefore, the harder a task is, the more one will become frustrated and switch to another task.

H5. When multitasking at their discretion, those who perceive the primary task as difficult will have more switches than those who consider the primary task not as difficult.

Table 5-1 Summarizes the Hypotheses.

Table 5-1. Summary of Hypotheses[4]

H #	*Hypothesis*
H1	Mandatory interruptions will have a negative effect on accuracy.
H1a	Those who multitask at their discretion will have greater accuracy than those forced to multitask.
H1b	Those who do not multitask will have greater accuracy than those forced to multitask.
H2	Those who do not multitask will complete their primary task faster than those who multitask.
H2a	Those who do not multitask will complete their primary task faster than those who are forced to multitask.
H2b	Those who are not multitasking will complete the primary task faster than those who can multitask at their discretion.
H3	When multitasking at their discretion, those who decide to switch tasks more often will have lower accuracy than those who switch less often.
H4	Those forced to multitask during a task they consider harder will perform worse than if they considered the task easier.
H5	When multitasking at their discretion, those who perceive the primary task as difficult will have more switches than those who consider the primary task not as difficult.

[4] This hypotheses table will be modified in the next chapter.

CHAPTER 6. METHODOLOGY

6.1 Pilot Design

In the Spring of 2009, a pilot study was conducted to test the procedures and the experimental conditions before running the actual experiment. In order to study the effects of multitasking on performance, three game-like tasks, all with objectively correct answers, were used.

6.1.1 Tasks

Each subject performed three different problem-solving game-like tasks (numeric, textual and visual), which needed to be completed in a given amount of time.

- The numeric task was a Sudoku puzzle.[5] The goal of a Sudoku puzzle is to fill in all the boxes in a 9 X 9 grid, so that each column, row, and 3 X 3 box have the numbers 1 – 9 without any of those numbers being repeated.
- The textual task consisted of unscrambling a series of letters to find as many words as possible using all or some of the letters given.
- The visual task required subjects to select the shape that best fit the pattern. Subjects were shown four shapes and had to choose the shape that did not belong.[6]

Sudoku was chosen as the primary task since it requires more mental concentration than the secondary tasks. When a subject returns to this primary task they need to remember their

---

[5] The Sudoku puzzle was taken from http://puzzles.about.com/od/sudokupuzzles/qt/Sudoku-Puzzles-To-Print.htm.

[6] Taken from http://www.mathworksheetscenter.com/mathskills/shapes/shapenotbelong/shapenotbelongK2begles.pdf

thought process. The word-puzzle also requires a bit more mental energy than the visual tasks, but less than the Sudoku puzzle. The goal was to implement tasks of different skills and duration that could be combined in different ways to emulate what happens in an actual computer usage session. Generally, users work on a primary task which requires concentration. They might be interrupted by an instant message alert, which will not require much concentration to respond to. Or, perhaps they have an email message that requires a little more concentration than an IM alert, but less than their primary task.

In order to prevent subjects from informing future participants of the answers to the tasks, for the visual task, the shapes were randomly placed in different orders for each subject. For the textual task, some subjects were shown the word "TEAR", while others were randomly shown "TARE" or "RATE". This way the words all had the same number of possibilities, yet they appeared to be different than each other.

6.1.2 Conditions

The initial design consisted of three different approaches for completing the above three tasks (discretionary multitasking, forced multitasking, or non-multitasking) and subjects were randomly assigned into one of the three.

- *Discretionary Multitasking (DMT):* All the tasks were presented at once, in different tabs and subjects were able to choose when to complete each task. This condition allowed subjects to switch tasks at any point. The interface kept track of when subjects were switching and how often.
- *Forced Multitasking (FMT)*: The secondary tasks appeared while the subjects were in the middle of completing the first task. In this condition, subjects were interrupted at

different intervals of time with pop-up windows that forced them to complete other tasks. The interrupting intervals were as follows: the second task appeared four minutes after the start of the exercise, and the third task after ten minutes. The interrupting task had to be completed before the user could resume the primary task. If the primary task was finished before receiving an interrupting task, the pending task was displayed upon completion of the primary task.

- *Non-Multitasking (NMT)*: In this condition, the second task was displayed as a pop-up window only after the first task was completed. Likewise, the third task was displayed after the second task was completed. If time ran out, i.e. the subject reached the 45 minute time limit, any task not finished remained uncompleted.

Microsoft Visual C++ 2005 was used in order to create the task-based interface as well as the pre-test and post-test questionnaires. Sample screenshots of the three conditions are shown in Appendix C.

6.1.3 Pilot Sample Subjects

For the pilot study, 110 Baruch College students were recruited through advertising in classrooms. They all signed an IRB-approved recruitment statement (See Appendix D). Subjects performed the computer-based experiment in a laboratory setting. There were 37 subjects in the discretionary multitasking condition, 36 in the forced multitasking condition, and 37 who were not allowed to multitask. All participants were offered a ten-dollar incentive for their participation.

6.1.4 Measures

Objective and subjective measures were collected. Objective measures consisted of actual performance indicators such as accuracy and completion time. Accuracy was calculated for each task. In Sudoku there were 45 empty spaces which the subjects needed to complete correctly. Percent correct is the number of squares entered correctly divided by the total number of empty squares (45). For the word task, there were ten words that the subjects were able to enter which can be made from unscrambling the letters. The percent correct is the number of correct words divided by 10.[7] For the visual task, this is a binary answer, yes or no. There was only one correct answer. A total score was calculated as an average of all three scores.

Subjective measures include perceptions about the experience included in the post-test questionnaire. In addition, a number of control variables such as demographics and their perception of their computer skills, Sudoku skills, and multitasking capabilities were also collected.

6.1.5 Questionnaire

Two questionnaires were included in this research. One was administered before the task was completed (pre-test questionnaire) and another one at the end of the exercise (post-test questionnaire).

---

[7] Words were verified using a website (http://www.wordplays.com/fcgi-bin/jumble.pl) which lists all possible combinations of words. Subjects listed up to ten 3 or 4 letter words that could be made out of the provided 4 letter word and their responses were compared with the answers provided by the above URL in order to determine the correct score.

Pre-test questionnaire: The pre-test questionnaire was designed to collect information to be used as controls. These include demographics, usage of and comfort with a computer, and prior experience with Sudoku. Demographics include age, gender, and academic level. They were also asked for their propensity to multitask with computer-based tasks in general. The questions for multitasking propensity were adapted from the Bluedorn et al. (1992) scale originally developed to measure monochronic or polychronic time use. In the modified scale, subjects were asked to rate, based on a 5-point Likert scale anchored by "strongly disagree" and "strongly agree", the following statements:

When I use my computer,

- I like to juggle several unrelated activities at the same time.
- I try to do many computer-based tasks at once.
- I work on one computer-based task at a time.
- I am comfortable carrying out several computer-based tasks at the same time.

Post-test questionnaire: The post-test questionnaire was designed to capture the experiences of the subjects when interacting with the gaming environment in their assigned condition and solving the tasks. The perception of the extent to which they multitasked in this session was measured using another adaptation of the Bluedorn et al. (1992) scale:

In this exercise,

- I switched between the assigned computer-based tasks.
- I tried to complete the assigned computer-based tasks at the same time.
- I worked on one computer-based task at a time.
- I was carrying out several computer-based tasks at the same time.

This perception provides a potential manipulation check of the integrity of the experimental conditions, and the extent to which subjects thought they multitasked in the discretionary (free switching) environment.

The post-test questionnaire also included a scale for perceived performance, i.e. how well they thought they performed. For this measure, an adaptation of Staples et al. (1999) scale was used. Subjects were asked to rate, based on a 5-point Likert scale with "strongly disagree" to "strongly agree", the following statements:

- I believe my work was effective.
- I would rate my performance in the top quarter.
- I am happy with the quality of my work output.
- I was able to work very efficiently.
- I was highly productive.
- I believe the way I worked was efficient.

The questions on their perception of their tasks performance and their multitasking are all mixed and listed in the post-test questionnaire in a random order. Additional questions were included to measure their perception of the level of complexity of the tasks, in order to see whether the level of difficulty of a task had a different impact on their performance in these multitasking conditions.

See Appendix E for screenshots of the pre- and post-test questionnaire.

6.1.6 Procedure

Subjects signed up to participate in an experimental session and were given a reminder message via email. Upon arrival, each subject was handed a general instruction sheet (see Appendix F) and a unique account number. The general instruction handout explained the three tasks. Instructions were also provided on the interface itself. After signing the consent form (Appendix D), subjects were set up into one of the three randomly assigned conditions. On the screen, subjects were then given more specific instructions pertaining to their environment as well as a pre-test questionnaire where they entered their account number. When the subject finished s/he was allowed to begin the tasks. Subjects were allowed at most 45 minutes to complete all of the tasks. After completing the tasks, or exceeding the 45 minute time limit, the subject was then directed to the post-test questionnaire.

The experimental sessions took place in a computer lab. Upon signing in, each subject was assigned to a computer. In each computer, the results of the tasks and questionnaires were automatically written to a unique log file. Upon finishing, subjects were compensated for their participation and their log files were copied from the local hard drives. A PERL program was then used to combine the results of all the log files to a single spreadsheet for data analysis.

6.1.7 Analysis

Once all the data were collected, performance was compared across all the conditions to determine the effects of the independent variables on the dependent variables. SAS was used to conduct ANOVAs in order to compare the performance in the three different conditions.

6.2 Pilot Results

Table 6-1 shows the breakdown of subjects per condition for the pilot experiment. As stated above, the distribution of subjects across conditions is 37, 37 and 36 for discretionary multitaskers (DMT), non-multitaskers (NMT) and forced multitaskers (FMT), respectively. As shown in Table 6-1, one user in the NMT condition did not have enough time to attempt either of the secondary tasks because the subject had exceeded the 45 minute time limit.

Table 6-1. Breakdown of Subjects Across Conditions

	Sudoku	Word	Visual
DMT	37	37	37
NMT	37	36	36
FMT	36	36	36

In the forced multitasking condition, 10 out of 36 subjects completed the third task, the visual exercise, without multitasking (i.e. in a sequential fashion) since they completed Sudoku before the time for the next interruption had occurred and therefore it was displayed upon the completion of their primary task. One subject never even got to the word task since the subject finished the Sudoku puzzle too quickly and therefore ended up completing all of the tasks in a sequential fashion.

Most of the results of the pilot analyses were not significant and one contributing factor could be due to the sample size which is not that large. There were only 37 or 36 subjects in each condition. However, in addition to increasing the sample size in the actual experiment, design modifications were also made based upon the results of the pilot experimentation. These modifications will be discussed in the next section.

The first hypothesis predicted that subjects forced to multitask would have the lowest accuracy. While forced multitaskers appeared to have the lowest accuracy for the primary task, the differences were not significant and therefore H1 was not supported (see Table 6-2). Total score is calculated as the mean of all the different scores.

Table 6-2. Means of Different Scores

Scores				
	Total	Sudoku	Word	Visual
DMT	80.91	81.92	87.84	72.97
NMT	77.25	80.24	84.17	66.67
FMT	79.90	70.25	86.11	83.33
F Value	0.30 ns	1.46 ns	0.36 ns	1.33 ns

The second hypothesis predicted that subjects who did not multitask would perform tasks faster than subjects in the other conditions since less multitasking would minimize the cognitive cost associated with switching. In terms of overall time, subjects tended to spend the shortest amount of time in the non-multitasking condition (and the most time in the discretionary condition), but the means were not significant and therefore H2 was not supported. A notable exception is the visual task where subjects in the discretionary multitasking condition spent significantly more time in the visual task than the forced multitataskers and non-multitaskers (see Table 6-3). This is probably the result of the way the game environment is designed because discretionary subjects can return to a task repeatedly while subjects in the other conditions cannot return to a previous task.

Table 6-3. Average Time for Tasks

Average Time (in minutes)				
	Total	Sudoku	Word	Visual
DMT	20.42	14.96	4.92	0.54
FMT	19.06	14.66	4.1	0.31
NMT	17.95	14.05	3.7	0.3
F Value	0.51 ns	0.08 ns	1.60 ns	10.59*

* p<.0001

Overall, subjects who never played Sudoku before performed significantly worse on Sudoku, regardless of their experimental condition (sud_score_{played} = 91.40 vs. $sud\_score_{neverplayed}$ = 44.38, F Value=90.54, p<.0001), and their total times were longer than those who had previously played tot_time_{played} = 22.33 vs. $tot\_time_{neverplayed}$ = 17.84, F Value=4.28, p=0.04. Out of the 110 subjects, 78 subjects had previously played Sudoku while 32 did not. After separating the results between those who never played and those who played, the non-multitasking condition appeared to be the best method for those who never played Sudoku and the forced multitasking condition the worst (although this was not significant). For those who have played before, the discretionary multitasking condition and forced multitasking condition appear to be better methods while the non-multitasking condition is the worst (See Table 6-4). For those who played Sudoku before, it is significant that the total score for subjects in the forced multitasking condition is higher than the total score for subjects who did not multitask (tot_score_{FMT}=89.73 $totscore_{NMT}$=77.25, F=6.50, p=.01). For the word task, subjects in the discretionary multitasking condition performed significantly better than subjects who did not multitask (wrd_score_{DMT}=90.71 wrd_score_{NMT}=80.80, F=4.15, p=.0467) and for the visual task the subjects in the forced multitasking condition performed significantly better than subjects in both the discretionary multitasking condition and non-multitasking condition as shown in Table 6-4.

Table 6-4. Accuracy for subjects who Played Sudoku Before vs. Never Played

Scores - Never Played				
	Total	Sudoku	Word	Visual
NMT	75.74	52.22	95	80
FMT	62.51	39.83	86.15	61.54
DMT	62.59	42.22	78.89	66.67
F Value	1.58 ns	0.53 ns	2.24 ns	0.43 ns
Scores - Played				
	Total	Sudoku	Word	Visual
NMT	77.84	90.62	80	61.54
FMT	89.73	87.44	86.09	95.65
DMT	86.8	94.68	90.71	75
F Value	2.91 ns	0.78 ns	2.23 ns	4.29*

* p<.05

In terms of the timing for subjects who never played Sudoku, non-multitaskers took the shortest amount of time, however it was not significant ($total\_time_{NMT}$=15.93 $total\_time_{FMT}$=25.72 $total\_time_{DMT}$=24.57, F=2.17). For the visual task, discretionary multitaskers took significantly longer than subjects in the forced multitasking condition (vis_time_{DMT}=0.67 vis_time_{FMT}=0.37, F=4.48, p=.0469). With regard to subjects who played Sudoku before, those in the forced multitasking condition had the shortest amount of time (though not significantly, $total\_time_{FMT}$=15.30 $total\_time_{NMT}$=18.69 $total\_time_{DMT}$=19.09, F=1.18). In the visual task discretionary multitaskers took significantly longer than both the forced multitaskers and non-multitaskers (vis_time_{DMT}=0.50 $total\_time_{NMT}$=0.26 $total\_time_{FMT}$=0.27, F=10.67, p<.0001).

The pilot analysis has shown that there is a difference in both accuracy and speed for those who played and never played Sudoku. This factor contributes to a later decision to add a practice Sudoku round for participants, as well as a control for the level of Sudoku skills of the participants when performing statistical analyses.

Hypothesis 3 predicted that subjects who were in the discretionary multitasking condition and switched more often would have lower accuracy than those who switched less often. This hypothesis was not supported in the pilot experiment.

Hypothesis 4 predicted that subjects who were forced to multitask and perceived a task as harder would have a lower accuracy than subjects who found it easier. As shown in Table 6-5 when examining the total scores, H4 is supported ($tot\_score_{Hard}=58.79$ $tot\_score_{Medium}=85.36$ $tot\_score_{Easy}=93.33$, $F=11.59$, $p=0.0002$). Those who found Sudoku easy in the forced multitasking condition did significantly better on Sudoku than those who found it a hard or medium level of difficulty. This was not true for non-multitaskers or discretionary multitaskers whose total scores were not significantly different for the level of difficulty. Tables 6-6 and 6-7 show the results when examining on the accuracy for the Sudoku task or the accuracy for the secondary tasks (visual and word). In fact, subjects who were forced to multitask and considered Sudoku to be difficult had a lower accuracy in the secondary tasks as well (see Table 6-7). This was not true for the discretionary multitasking or non-multitasking conditions.

When examining table 6-5 horizontally rather than vertically, while insignificant, it did appear that subjects in the forced multitasking condition who found Sudoku to be harder had a lower accuracy (58.79) than the other two conditions (79.05 in NMT and 73.54 in DMT). If this was supported, H1 would have been partially supported since the forced multitasking condition had the lower accuracy but only when subjects found the primary task difficult. This will be explored again in the analysis for the actual experiment (non-pilot). When subjects found Sudoku to be easier, it also appeared that these forced multitaskers performed better than the other two conditions. This pattern also held true when examining only the Sudoku scores and only the secondary task scores, yet they are also not significant (see Table 6-6 and 6-7).

Table 6-5. Total Accuracy and Level of Difficulty

	DMT	NMT	FMT		F Value
Easy	87.84	82.28	93.33		1.63 ns
Medium	79.33	72.38	85.36		1.54 ns
Hard	73.54	79.05	58.79		1.75 ns
F Value	1.64 ns	0.87 ns	11.59 *		

*p<0.001

Table 6-6. Sudoku Scores and Level of Difficulty

	DMT	NMT	FMT		F Value
Easy	93.5	83.76	100		1.63 ns
Medium	82.67	84.03	65.3		1.54 ns
Hard	63.95	66.94	43.64		1.75 ns
F Value	3.59*	1.06 ns	10.43**		

*p<.05 **P<.001

Table 6-7. Secondary Tasks' Scores and Level of Difficulty

	DMT	NMT	FMT		F Value
Easy	85	81.54	90		1.63 ns
Medium	77.67	66.56	95.39		1.54 ns
Hard	78.33	84.29	66.36		1.75 ns
F Value	0.33 ns	1.45 ns	7.87*		

*p<.01

The last and fifth hypothesis examined whether those who perceived the level of difficulty of Sudoku as difficult would have more switches (i.e. more multitasking) than those who found it to be easier. When examining only the discretionary multitaskers, those who found Sudoku difficult had significantly more switches than those who found it to be easier or a medium level of difficulty, the Duncan values are in parentheses: sud_diff_{Hard} =15.78 (A) sud_diff_{Medium} =9.47 (B) sud_diff_{Easy} =8.31 (B), F=3.83, p=.0316. Therefore, H5 was supported in this pilot.

6.3 Design Modifications

6.3.1 More Tasks

Analysis of pilot data showed that having an environment where subjects were forced to multitask, but only interrupted twice or sometimes only once if they finished early does not represent a realistic environment with several interruptions. Having only three tasks also limited the behavior of the subjects in other conditions, particularly in the discretionary multitasking condition. Therefore, a decision was made to expand the number of tasks. Instead of only two secondary tasks, the new environment provided five smaller tasks, similar to the ones used in the pilot test.

In addition to the primary task which was still Sudoku and the textual word-based task, there were now two visual-based tasks instead of one. Newly added tasks were numeric series problem-solving, where subjects had to guess the missing number in the series of numbers presented. There were two number series task sets.

6.3.2 Controlled Time

Given the different times that pilot subjects took to complete the tasks, a decision was made to include time controls. This restriction intends to rule out the effect of time on differences in performance. Accordingly, each task was timed with a time limit intentionally shorter than the amount of time a subject would need to complete the task. This decision seeks to avoid the following cases that occurred in the pilot:

1. *Forced Multitasking Condition (FMT)*: A subject in the forced multitasking condition was able to end up performing tasks sequentially (without multitasking) due to finishing the main task too early. Whether the subject ended the task too early because s/he was done early or because s/he was frustrated, this situation can now never occur. Time is now intentionally too short to complete a task. In addition, a subject cannot decide to end a task early, but rather has no choice but to continue even if s/he is frustrated. There is no longer the capability to end a task earlier than the allotted time period. Therefore, forced multitaskers will have to complete all secondary tasks as interrupting tasks at different intervals while completing the primary task (i.e. it can no longer occur sequentially).
2. *Non-Multitasking Condition:* We saw in the pilot that it is possible for some subjects to never reach any of the secondary tasks if they reached their 45 minute time limit during the primary task. Therefore, each task now has an individual time limit, so once the time limit for the primary task has been reached the system will automatically move the subject into the next task and all subjects have the same amount of time and cannot exceed any total time limits. The system has more control in this case and not the user.
3. *Discretionary Multitasking Condition:* A subject in the discretionary condition could in theory not participate in every task, similar to what we saw in the non-multitasking condition. A discretionary subject could stay on Sudoku the entire

time and choose to ignore the other tabs/tasks. In the new experiment design, discretionary subjects have to attempt every task and complete them before the program ends. In addition, we saw that the discretionary timing for the visual task was significantly greater than in the other conditions. Now, in the discretionary condition, although tasks can be returned to repeatedly, each task will have the same amount of time as in the other conditions.

The task times were chosen based on additional pilot testing to determine time limits that were shorter than average subjects would need to complete the tasks. This also prevented someone from being idle (since in that case they would not be multitasking). In addition, each task was made longer, as will be explained below, to ensure that subjects did not complete any task too early but rather were not able to ever finish a task.

Sudoku: As stated, the primary task was a Sudoku puzzle. However, the Sudoku puzzle was modified to a harder puzzle (additional pilot testing showed that people found it harder). The reason for this was to keep the subjects occupied longer (instead of 45 empty squares there were now 49). Time for this task was set to 18 minutes. And in order to ensure that subjects who did finish the puzzle early were not idle, a second optional Sudoku puzzle was provided for anyone who finished too soon.

Word: The Textual or Word task was modified to be a slightly longer task than the previous word problem and now had 20 empty boxes available instead of 10. The word was randomized to “DATER”, ”RATED”, or “TRADE” since they all have the same letters and therefore the same amount of words that can be created with those letters. Subjects were given 1.5 minutes to complete the task.

Visual: Instead of only one visual task being displayed. There were now ten visual multiple choice problems. Time was set to 48 seconds to ensure that no one could complete all ten problems in time. This was to mimic an infinite amount of questions. There were two of these visual tasks (i.e. two sets of ten visual exercises).[8]

Number Series: As described previously, a Number Series exercise was created. Subjects had two number series exercises to complete, each with ten questions.[9] The time for each number series task was set to 48 seconds.

Performance in the three different conditions could now be compared accurately since everyone in all conditions completed the same number of tasks and spent the same amount of time in each task. The total time for completing all tasks was 22.7 minutes (18 minutes + 1.5 minutes + 48 seconds * 4).

In addition, the new version of the experiment had a practice round of Sudoku. Many users commented in the post-test questionnaire of the pilot study that they felt that the Sudoku was very hard. And as we saw in the analysis of the pilot test data, those who had not played

---

[8] In addition to the previous URL mentioned for the visual exercise in the pilot, others were either made up or taken or modified from the following places: http://www.intelligencetest.com/questions/visualization.htm, http://www.iqtestexperts.com/visual-sample.php, http://www.didax.com/newsletter/pdfs/mental_math_215297.pdf, and http://www.syvum.com/iq/.

[9] Number Series task problems were taken from http://www.funbrain.com/cracker/index.html.

before did significantly worse than the others. In order to introduce subjects who had never played Sudoku before to the nature of the puzzle, everyone was given a practice Sudoku puzzle first to familiarize them with this type of problem. The subjects had up to 10 minutes to complete the Sudoku practice round. This practice round was optional. The subjects had the option to end the practice round early or to skip it entirely. The Sudoku practice-puzzle was very simple and generated popup windows any time an incorrect number was entered. These popup windows explained why the number entered was incorrect.

6.3.3 Task Randomization

In all conditions, Sudoku was always the primary task. It was presented first in the forced and non-multitasking conditions, and it appeared in the first tab in the discretionary treatment. However, to avoid order effects in overall performance measures, the order of the secondary tasks, in all conditions, was randomized. In the forced and non-multitasking conditions, a random number generator determined the order in which secondary tasks were going to appear. A similar approach was used to decide the order in which the tabs were going to be displayed in the discretionary condition.

Screenshots for the practice round as well as each task (in a discretionary environment) can be seen in Appendix G.

6.3.4. Interface

In addition to the tasks, time controls, and randomization, there were two other major changes to the design of the game environment: one in the instructions and another in terms of window size.

Instructions: Instead of the previous instruction handout, subjects were given different instruction handouts depending on their particular condition. Therefore, instructions did not need to be on the computer screen at all times, and subjects had the reference as a handout in front of them when needed. In addition, they were told to read it before beginning and there were instructions on the computer screen as well that appeared before they started the actual experiment. This was done to create more screen space for the actual tasks and to avoid subjects losing time and getting distracted by extra information on their screen.

Window Size: In the forced multitasking and non-multitasking conditions the popup windows with the secondary tasks were sized and positioned to block most of the primary task. This was mostly a concern for the forced condition, so that subjects would not get distracted from the secondary task.

See Appendix H for revised handout instructions, on screen instructions, and a screenshot of the larger window size.

6.3.5 Added and Modified Hypotheses

In the research questions and hypotheses chapters, performance is discussed in terms of accuracy and time. However, time is now set for each task and can therefore no longer be examined as a dependent variable. Since time was set intentionally shorter than the average amount of time a person would need to complete each task, performance can be measured in terms of productivity, i.e. how much of the tasks subjects were able to complete. Since non-multitaskers do not have to spend their time recalling previous tasks, therefore it is predicted that they will perform better in terms of productivity than the other two conditions.

Hypothesis 2 can therefore be modified accordingly:

Previous Hypothesis 2:

H2. Those who do not multitask will complete their primary task faster than those who multitask.

H2a. Those who do not multitask will complete the primary task faster than those who are forced to multitask.

H2b. Those who do not multitask will complete the primary task faster than those who multitask at their discretion.

Updated Hypothesis 2:

H2. Those who do not multitask will have greater productivity than those who multitask.

H2a. Those who do not multitask will have greater productivity than those forced to multitask.

H2b. Those who do not multitask will have greater productivity than those who multitask at their discretion.

When discussing *performance,* the term will refer to both *accuracy* and *productivity.* While *productivity* is defined as how much of the tasks subjects were able to complete, *accuracy* is dependent on how much subjects were able to perform correctly.

Since the results of the pilot testing showed no significant differences in terms of the scores, it is possible that negative perceptions about the task or the exercise such as anxiety, frustration, or annoyance would be different among conditions. Therefore a sixth hypothesis was added.

Being interrupted can lead to more anxiety and annoyance (Bailey and Konstan 2006; Mark et al. 2008). In addition, interruptions can lead to more frustration (Iqbal and Bailey 2008; Mark et al. 2008). While performing a task, receiving interruptions can cause a user to feel pressured and become more frustrated than they normally would be. Having additional tasks interrupting a user while s/he is performing a primary task can lead to more anxiety, annoyance, and frustration since the individual may have a deadline with which to finish the main task, and becoming interrupted with a second task can cause additional time pressure.

H6. Receiving external interruptions will lead to more anxiety, annoyance, and frustration.

H6a. Those forced to multitask will have more anxiety, annoyance, and frustration than those who do not multitask.

H6b. Those forced to multitask will have more anxiety, annoyance, and frustration than those who multitask at their discretion.

Table 6-8. Revised Summary of Hypotheses

H #	*Hypothesis*
H1	Mandatory interruptions will have a negative effect on accuracy.
H1a	Those who multitask at their discretion will have greater accuracy than those forced to multitask.
H1b	Those who do not multitask will have greater accuracy than those forced to multitask.
H2	Those who do not multitask will have greater productivity than those who multitask.
H2a	Those who do not multitask will have greater productivity than those forced to multitask.
H2b	Those who do not multitask will have greater productivity than those who multitask at their discretion.
H3	When multitasking at their discretion, those who decide to switch more often will have lower accuracy than those who switch less often.
H4	Those forced to multitask during a task that they consider harder will perform worse than if they considered the task easier.
H5	When multitasking at their discretion, those who perceive the primary task as difficult will have more switches than those who consider the primary task not as difficult.
H6	Receiving external interruptions will lead to more anxiety, annoyance, and frustration.
H6a	Those forced to multitask will have more anxiety, annoyance, and frustration than those who do not multitask.
H6b	Those forced to multitask will have more anxiety, annoyance, and frustration than those who multitask at their discretion.

6.3.6 Expansions to the Post-Test Questionnaire

Although task perceptions had not been considered for the pilot testing, perceptual measures were added to determine the users' level of anxiety, annoyance, and frustration during each condition.

Anxiety was measured using a slightly modified version of three out of the four scales that Thatcher and Perrewe (2002) used to measure computer anxiety:[10]

- I felt apprehensive about using this system.
- I hesitated to use this system for fear of making mistakes that I could not correct.
- This system was somewhat intimidating to me.

Annoyance was measured using a modified version of the two annoyance statements from Sanderman et al. (1998):

- The system was annoying.
- It was annoying that the system made assumptions on how I like to solve multiple tasks.

Frustration with the system, i.e. the environmental condition, was measured using a slightly modified form of three of the five scales Ceaparu et al. (2004) used to measure frustration:[11]

- Overall, I was frustrated with this system.
- These frustrations will affect me the rest of the day.
- These frustrating experiences impacted my ability to get the tasks done.

---

[10] One item addressed computer anxiety over the computer destroying large amounts of information which was not relevant to our interface design and was therefore not included.

[11] Two items were not included: one dealing with the interaction with co-workers and one asking about typical computer frustration.

Another modification to the post-test questionnaire was that instead of using easy, medium, and hard for the difficulty levels of the tasks, we modified it consistent to how Bailey and Konstan (2006) measure difficulty, except ranging from 1 to 5 instead of 1 to 6 to keep it consistent with our 5 point-scale for the rest of the questions.

An open-ended question asking why subjects switched was added after the post-test questionnaire only for discretionary subjects in order to learn more about why people are multitasking. See Appendix I for an updated post-test questionnaire as well as a screenshot for the open-ended discretionary question.

CHAPTER 7. RESULTS

7.1 Description of the Sample

7.1.1 Final Sample

A total of 636 subjects participated in the experiment. This sample does not include those who took part in the pilot study. Although they all used the revised interface and modified tasks, they were recruited with two different methods. About half of the sample (307 subjects) received monetary compensation ($10) for their participation, while the rest (329 subjects) were recruited later from the SCIS Department Subject Pool and received course credit for their participation. The revised IRB consent form for those receiving credit can be found in Appendix J.

7.1.2 Demographic Variables

The 636 subjects were randomly assigned to each condition. The final distribution is as follows: 212 subjects in NMT (non-multitasking condition), 212 in FMT (forced multitasking condition), and 212 in DMT (discretionary multitasking condition). Demographic characteristics gathered in the pre-test questionnaire include age, gender, academic level, computer skills, and previous level of Sudoku experience. The descriptive statistics of the continuous variables (age, computer skills, and Sudoku experience) are presented in Table 7-1 for all 636 subjects.

Table 7-1. Descriptive Statistics (N=636)

	Mean	S.D.	Min	Max
Pre-test Variables				
Age[1]	22.44	4.69	18	54
Computer Skills[2]	3.70	0.80	1	5
Sudoku Experience[3]	1.55	1.43	0	5

1 One subject typed an invalid age and was omitted from this analysis.
2 Computer skills was measured with a 5 point scale from 1-Poor to 5-Excellent.
3 Sudoku Experience was measured with a 0 to 5 scale similar to computer skills but with a 0 for those who had never played Sudoku before.

The average age of the entire sample is 22.44 years, the average level of computer experience is between average and good (3.7) and the prior level of Sudoku skills is rather low (1.55 out of 5) because 240 participants (out of 636) had never played Sudoku before. With regard to gender, 334 subjects were male (53%) and 302 were female (47%). The breakdown of the sample in terms of the academic level is as follows: 14 were freshman (2%), 17 graduate (3%), and 73 seniors (11%). The majority of subjects were at the middle undergraduate levels: 271 participants (43%) were sophomores and 261 (41%) were juniors.

Two additional questions were included in the pre-test questionnaire. First, there was a yes/no question on whether the subject had a computer at home. Only 3 subjects (less than 1%) did not have a computer at home, while 636 did. Second, there was a scale to investigate the multitasking propensity of the participant. This scale will be described in a future section.

To ensure that randomization worked and to rule out alternative explanations, the pre-existing differences among participants were first checked against the different conditions. None of the continuous pre-test questionnaire variables showed a systematic variation. Separate ANOVAs were performed using age ($Mean_{DMT}$= 21.98; $Mean_{NMT}$ = 22.53; $Mean_{FMT}$ = 22.80; $F(2,632)$ =1.71 ns), computer skills ($Mean_{DMT}$ = 3.71; $Mean_{NMT}$ = 3.70; $Mean_{FMT}$ = 3.70; $F(2,633)$ =0.02 ns), and Sudoku experience ($Mean_{DMT}$ = 1.54; $Mean_{NMT}$ = 1.57; $Mean_{FMT}$ = 1.53; $F(2,633)$ =0.04 ns), as dependent variables. There was no significance across conditions for any of these variables indicating that there were no systematic variations within conditions on any of these variables. Separate chi-square analyses were conducted for gender and for academic level. For gender, the results showed that male and female participants were equally distributed across conditions (χ^2 = 1.68; p=0.43 ns), and similar results were obtained for academic level (χ^2 = 4.63; p=0.80 ns).

Since about half of the subjects (307) received \$10 for their participation and the other half (329 subjects) received course credit, a chi-square analysis was also done for the paid vs. credit groups. The results show that the two sub-samples (paid and credit) were equally distributed across conditions ($\chi^2 = 0.01$; p=0.99 ns).

Although these demographic variables (age, gender, academic level, computer skills, Sudoku experience), and the reward method (paid vs. credit) were all randomly distributed into each condition, they will be used as controls throughout the analysis of the experiment.

7.1.3 Paid vs. Credit

As explained above about half of the subjects received monetary compensation for their time, while half received course credit. Students in the Introduction to Information Systems course at Baruch College were required to be a participant in a study and were provided with different studies they could choose from.

Separate ANOVAs were performed comparing the paid and credit groups and the demographic variables. There is a significant difference in terms of the age of the two groups ($Mean_{Credit}$= 23.23; $Mean_{Paid}$ = 21.59; F(1,633) =20.01, p< .0001). The paid group had a lower average age. While there are older students in the course who had to choose a study to be a participant in, this was not the case for those who wanted to participate in the experiment for their \$10 incentive. For example, there were two students over 50 and seven over 40 in the credit group, while the oldest age in the paid group was 35. Subjects who were paid also had better computer skills ($Mean_{Paid}$= 3.82; $Mean_{Credit}$ = 3.59; F(1,634) =13.02, p=0.0003), and while their Sudoku experience was better it was not significant ($Mean_{Paid}$= 1.66; $Mean_{Credit}$ = 1.44; F(1,634) =3.69 ns). Chi-square analyses were conducted for gender and it was significant (χ^2 =

6.29; p=0.01). See Table 7-2 for the breakdown of gender and paid vs. credit. The academic level was also significant (χ^2 = 58.30; p<.0001) as shown in Table 7-3.

Table 7-2. Distribution of Males and Females in the Paid and Credit Groups

	MALE	FEMALE
CREDIT	157	172
PAID	177	130

Table 7-3. Distribution of Academic Level in the Paid and Credit Groups

	FRESHMAN	SOPHOMORE	JUNIOR	SENIOR	GRADUATE
CREDIT	6	156	153	14	0
PAID	8	115	108	59	17

The above table shows that there were some graduate students who participated from the paid group, since subjects were recruited from many different courses. However, the credit group is more homogenous in that it only consisted of subjects enrolled in the course, CIS 2200, Introduction to Information Systems.

While there were differences in terms of the demographic variables in the paid vs. credit sample, as mentioned in the previous section subjects who were paid and subjects who received course credit were evenly distributed across conditions and this variable (whether or not the subject received compensation) will be used as a control throughout the analyses.

7.1.4 Pre-test Variables

In the pre-test questionnaire, subjects' propensity to multitask was determined via four questions adapted from Bluedorn et al. (1992), each with a 5-point Likert scale, on whether they tended to multitask in general on the computer. One question worded in the opposite direction was reversed to have consistent responses with the other three questions. A confirmatory factor analysis of these items showed a single factor with items loadings of .55 or higher (See Table 7-4). The cut-off of .55 was chosen following the recommendation of Hair et al. (1998) on p. 112.

The scale with the four items had a high level of reliability (Cronbach's Raw Alpha=.743, Standardized Alpha = .750). A composite variable, multitasking propensity, was created by averaging these four items for each participant. As in the case of the other pre-test variables, we tested whether there was a systematic variation of this variable across conditions with an ANOVA analysis, using multitasking propensity as the dependent variable. The results show that it did not vary across conditions ($mt\_propensity_{DMT}=3.61$; $mt\_propensity_{NMT}=3.52$; $mt\_propensity_{FMT}=3.50$; $F(2,633)=1.43$ ns).

Table 7-4. Factor Analysis of Multitasking Propensity

Item	*Factor 1*
I am comfortable carrying out several computer-based tasks at the same time.	0.82
I try to do many computer-based tasks at once.	0.80
I like to juggle several unrelated activities at the same time.	0.79
I work on one computer-based task at a time.	0.61

7.2 Measures of Post-Test Variables

After the participant completed the experiment by solving the tasks in the assigned condition, a post-test questionnaire was presented to collect perceptions of the process.

7.2.1 Subjective Measures

Subjective measures collected in the post-test questionnaire include: subjects' perceptions on how they multitasked in the experiment (4 items adapted from Bluedorn et al. (1992)), their perception of their performance (6 items adapted from Staples et al. (1999)), and level of anxiety, frustration, and annoyance (eight questions: three for frustration from Ceaparu et al. (2004), three for anxiety from Thatcher and Perrewe (2002), and two for annoyance from Sanderman et al. (1998)).

A factor analysis showed three factors (see Table 7-5).

Table 7-5. Factor Analysis of Post-Test Data

Item	*Factor 1*	*Factor 2*	*Factor 3*
I am happy with the quality of my work output.	**0.79323**	-0.10785	0.05237
I believe the way I worked was efficient.	**0.76776**	-0.20029	0.13187
I believe my work was effective.	**0.76284**	-0.07062	-0.03735
I was highly productive.	**0.7437**	-0.12883	0.17116
I would rate my performance in the top quarter.	**0.72817**	0.06957	0.12357
I was able to work very efficiently.	**0.71563**	-0.17251	0.13263
This system was somewhat intimidating to me.	-0.03523	**0.7387**	-0.02592
These frustrating experiences impacted my ability to get the tasks done.	-0.10814	**0.7308**	-0.0655
Overall, I was frustrated with this system.	-0.31485	**0.67224**	0.00202
It was annoying that the system made assumptions on how I like to solve multiple tasks.	-0.01053	**0.66609**	-0.03894
The system was annoying.	-0.22275	**0.66138**	-0.03369
These frustrations will affect me the rest of the day.	0.00069	**0.58641**	-0.02164
I hesitated to use this system for fear of making mistakes that I could not correct.	-0.10415	**0.56852**	0.13001
I felt apprehensive about using this system.**	0.12105	**0.2428**	0.20931
I was carrying out several computer-based tasks at the same time.	0.21536	0.00772	**0.76971**
I switched between the assigned computer-based tasks.	0.10858	-0.0525	**0.76012**
I worked on one computer-based task at a time.*	-0.13633	-0.1115	**0.72768**
I tried to complete the assigned computer-based tasks at the same time.	0.23337	0.12577	**0.66334**

* Item reversed; ** Item dropped

As shown in the above table, all the items for the perception of the subject's performance (the first six questions) loaded together with loadings greater than .70 (well above the .55 cutoff). These six items, were averaged together to form a performance perception composite variable or index. The last four items in the table, measured the perception of multitasking during the experiment. Note that one of the items in multitasking perceptions was reverse coded. Item loadings were all greater than .55 as well and were therefore averaged together to create a multitasking perception composite variable.

For anxiety, annoyance, and frustration, the factor analysis showed all eight questions loaded in a single factor (i.e. the second factor in the table above). Frustration and anxiety and annoyance can be perceived from a subject's perspective as very similar feelings. Therefore, these items were grouped together as the same category. All items loaded above .55, except for one item which was removed from the scale because its loading was very low. The remaining seven items were averaged together to create a composite variable that will be called heretofore frustration index.

The means, standard deviations, min, max, and the Cronbach Alpha, both raw and standardized, of the new variables can be seen on Table 7-6. All indexes had a high level of reliability.

Table 7-6. Descriptive Statistics of Post-Test Variables (N=636)

	Mean	**S.D.**	**Min**	**Max**	**Cronbach Alpha (Raw)**	**Cronbach Alpha (Standardized)**
Post-test Variables						
Performance Perception	2.75	0.73	1	5	0.862	0.862
Multitasking Perception	2.84	0.79	1	5	0.729	0.730
Frustration	2.64	0.69	1	5	0.800	0.801

In order to first check whether the subjects perceived the conditions as programmed in the game environment, the multitasking perception scale was used to perform a manipulation check. The mean of the multitasking perception in the forced multitasking condition was significantly higher than those in the other conditions, see Duncan values in the parentheses ($mtperc_{FMT}$= 3.17 (A) $mtperc_{DMT}$ = 2.72 (B) $mtperc_{NMT}$= 2.62 (B); $F(2,633) = 31.75$, $p<.0001$). This test provides evidence of the integrity of the conditions as implemented in the custom-developed environment. These results indicate that users who were forced to multitask perceived the environment as involving more multitasking than those in the other two conditions.

According to the Duncan groupings, subjects in the discretionary and non-multitasking condition reported similar perception of multitasking. However, 46 subjects originally assigned to the discretionary multitasking condition chose not to multitask, and worked in a sequential fashion. When those subjects are grouped with the non-multitaskers, then the perceptions of multitasking in the discretionary condition are significantly greater than in the non-multitasking condition, while the perceptions in the forced multitasking condition are still the highest ($Mean_{FMT}$= 3.17 (A) $Mean_{DMT}$ = 2.89 (B) $Mean_{NMT}$= 2.53 (C); $F(2,633) = 44.16$, $p<.0001$). This implies that the subjects perceived the multitasking environments as intended. Those subjects in the forced multitasking condition perceived that they performed the most multitasking, next was the discretionary group, and lastly the non-multitaskers. Because of this pattern, when examining multitasking performance across conditions, those in the discretionary group who worked sequentially (i.e. did not multitask) were moved into the non-multitasking group. Therefore, for the analysis 166 subjects are considered in the discretionary multitasking condition (212 minus 46).

The means for the subjects' perception of their performance was not significantly different across conditions ($Mean_{NMT}$= 2.76 $Mean_{DMT}$ = 2.76 $Mean_{FMT}$= 2.72; $F(2,633) = 0.26$ ns). Therefore, this variable will not be used to test the hypotheses. It is worth noting that the correlation between perception of performance and actual performance (measured by total accuracy score) is 0.296 ($p<.0001$).

The differences across conditions of the frustration variable will be examined in the Test of Hypotheses section.

The post-test questionnaire also asked subjects to rank the level of difficulty of the different tasks (where 1 is easy and 5 is hard). Table 7-7 shows the means and standard deviations. There were two Sudoku puzzles and not all subjects attempted the second Sudoku puzzle, therefore they were allowed to fill in N/A if they did not attempt the second puzzle. Also note that while there were two visual tasks, there was only one level of difficulty question for the visual task since subjects would probably not remember the difference they felt in difficulty between the first and second visual task and since these were randomized, subjects encountered them in a different order. The same held true for the number series task. Thus, only one measure of difficulty was collected for each type of task, except for Sudoku.

Table 7-7. Descriptive Statistics of Level of Difficulty Questions

	Mean	**S.D.**	**Min**	**Max**
Level of Difficulty				
Sudoku 1	3.22	1.38	1	5
Sudoku 2	3.80	1.03	1	5
Word	2.94	1.06	1	5
Visual	3.34	0.96	1	5
Number Series	3.08	1.05	1	5

As shown in the above table, subjects who wrote that they attempted the second Sudoku (360 subjects) did feel that it was harder than the first, as was intended. Based on activity data, only 252 subjects actually attempted the second Sudoku by entering some numbers. It is possible that the 108 subjects who indicated attempt, were thinking about it but never had a chance to enter any number. The purpose of the second puzzle was in order to ensure that subjects who finished the first were not idle, so a harder puzzle was chosen. The performance in the second Sudoku puzzle is not considered in the analysis of the results in future sections since not everyone attempted this second puzzle. Therefore, the score for this puzzle will not provide as much information as the analysis for all 636 subjects.

The difficulty level for Sudoku 1, word, visual, and number series did not vary across conditions. However, those in the forced multitasking condition who attempted the second Sudoku puzzle found it more difficult than those who did not multitask, see Table 7-8. This implies that for the second Sudoku puzzle the condition did affect how users ranked the level of difficulty. Being interrupted during a hard task caused subjects to consider the task even harder, as seen when subjects in the forced multitasking condition ranked the second Sudoku puzzle as harder than subjects who did not multitask. When examining only those subjects who wrote that they attempted the second Sudoku puzzle, there were 130 subjects in the forced multitasking condition, 93 in discretionary, and 137 in the non-multitasking condition. Note that the division across conditions is after the 46 discretionary multitaskers who worked sequentially were moved into the non-multitasking condition. Only 22 of the 46 subjects wrote that they attempted the second Sudoku puzzle. Therefore, the discretionary condition has fewer subjects than the non-multitasking condition in the table below.

Table 7-8. Comparison of Multitasking Conditions for Sudoku 2 Difficulty

Multitasking Category	*Sudoku 2 Level of Difficulty*
FMT	3.95 (A)
DMT	3.83 (A & B)
NMT	3.64 (B)
F Value (2, 357)	3.09*

Significance levels: * $p<.05$
Duncan grouping (A or B) in parentheses.

A chi-square analysis was done between those who attempted the second Sudoku puzzle and the condition. It is not significant ($\chi^2 = 3.23$; $p=0.20$ ns).

7.2.2 Objective Measures

In addition to the composite variables that were created from the post-test questionnaire, there were performance variables calculated from the subjects' activity during the actual experiment. Performance is conceptualized in terms of effectiveness and efficiency. For effectiveness, we use accuracy (number of correct answers), and for efficiency, we use productivity (number of answers provided regardless of whether they are correct).

The *accuracy* variable was calculated using the subject's scores in each task. The scores were calculated with the number of correct answers as a percentage of the total answers required. For example, in the Sudoku task there were 49 empty spaces that needed to be filled out with the appropriate numbers. The score was the number of correct values entered divided by the total number of empty squares. For the word task, there were 20 acceptable words that could be generated from unscrambling the letters. The percent correct is the number of correct responses out of 20. The same method was applied to calculate the visual and number series tasks' scores. The total score was computed by averaging the accuracy scores of all six tasks. Please note that total scores for both productivity and accuracy did not include the Sudoku 2 task because all 636 subjects would not have been able to be included in the analysis in that case.

The *productivity* measure was determined based on the amount of work completed per task, regardless of whether the answers were correct. Some subjects were able to complete more than other subjects. For example, the productivity measure was the percentage of the 49 empty Sudoku spaces that were filled out, or the number of words out of 20 that were typed into the word solution boxes. The total productivity score was computed by combining the percentage completion scores across the six tasks. The correlation between the accuracy and productivity variables is .701 ($p<.0001$).

Accuracy and productivity performance measures were automatically calculated by the custom-develop application log data generated for each user. The results can determine if there are any differences in performance for those in a forced multitasking condition vs. those in a discretionary multitasking condition or non-multitasking condition. Furthermore, an exploratory analysis within the discretionary condition was performed in order to test whether the accuracy or productivity performance varied depending on the number of switches. The means of total accuracy and total productivity for the entire sample can be found in Table 7-9.

Table 7-9. Means of Performance Variables

[N=636]	*Mean*	*Std Dev*	*Minimum*	*Maximum*
Accuracy	39.48	12.57	10.10	75
Productivity	58.64	13.26	19.30	93.33

Number of switches was calculated based on the number of tab clicks. The discretionary condition had six tabs, and each tab contained one of the six tasks. Therefore, the minimum number of switches a subject could have is five switches. This indicates that despite the flexibility afforded by this condition, some participants chose not to multitask. Since there were six tasks/tabs, the minimum number of tab changes in the discretionary condition was five. This occurs if a participant never returned to a previously used tab. Participants were able to switch tasks at any point by clicking on the corresponding tab. However, if the time limit on a tab was reached that tab became disabled and the user was unable to return to that task. All tabs had to be attempted at least once as the multitasking environment did not terminate until every task/tab's time expired. Clicking on the current tab did not increment the number of switches. The average number of switches in the discretionary multitasking condition (after the 46 non-multitaskers were removed) was 9.21 with a standard deviation of 4.37 (Min: 6; Max: 29), where N was 166.

In addition to the total number of switches, the system also kept track of the number of switches *to* each task. For example, the *Sudoku switches* variable is the number of tab clicks to Sudoku, and the *word switches* variable is the number of switches to the word task. This is calculated by counting the number of times the word tab was clicked on.

The number of switches *from* each task was also calculated. This is slightly different than the above. While the above examined exactly how many times a task was switched to, this kept track of how many times a task was switched out of. This was calculated by counting the number of times a subject clicked out of a task (i.e. when a subject clicked on a new tab, a counter for the current task was incremented).

7.3 Test of Hypotheses

One of the main goals of this study was to compare the performance between those who were able and chose to multitask (DMT), those who were forced to multitask (FMT), and those who did not multitask (NMT). Table 7-10 shows the number of subjects that are in each condition.

Table 7-10. Number of Subjects in Each Condition

Conditions	*Number of Subjects Included*
Forced Multitasking (FMT)	212
Discretionary Multitasking (DMT)	166*
Non-Multitasking (NMT)	258

*46 subjects from the DMT condition were moved into the NMT condition because they did not multitask.

7.3.1 Hypothesis 1

H1 predicted that the accuracy score for subjects in the forced condition would be lower than those in the discretionary or non-multitasking condition. When comparing the means of the accuracy performance across conditions, although the model is significant ($F(8, 626)=42.41$; $p<.0001$), there is no significance in terms of the multitasking mode (i.e. condition) as can be seen in Table 7-11. While discretionary multitaskers appear to have the highest accuracy and forced multitaskers the worst (as per H1), this difference is not significant. H1a predicted that those forced to multitask would have a lower accuracy than the discretionary multitaskers and H1b predicted that forced multitaskers would perform worse than non-multitaskers. Since the multitasking mode is not significant, neither H1a nor H1b is supported.

Table 7-11. Comparison of Multitasking Conditions for Accuracy

Multitasking Category	*Accuracy*
DMT	40.53 (A)
NMT	39.67 (A & B)
FMT	38.33 (B)
Model F(8, 626)	42.41 ***
R^2	35.15%
Explanatory Variables	
Condition	1.13 ns
Sudoku Experience	224.50***
Gender	25.48***
Age[1]	27.82***
Computer Skills	3.38 ns
Academic Level	0.03 ns
Paid (vs. credit)	0.54 ns

1 When age is used as a control, one non-multitasker was discarded since the subject entered an incorrect age (the subject entered “3”, a typo).

Significance levels: * p<.05; ** p<.01; *** p<.001;
Duncan grouping (A or B) in parentheses.

When comparing performance across conditions for only those subjects who found Sudoku more difficult, the subjects in the forced multitasking condition do perform significantly worse in terms of accuracy than the other two conditions. The mean for the Sudoku difficulty variable is 3.22 and the median is 3. Therefore, subjects who reported their level of difficulty for Sudoku as either a 1 or a 2, i.e. those who felt the task was easy, were discarded. Those who felt the task was more difficult (their Sudoku level of difficulty was reported as either a 3, 4, or 5) remained. N in this case was 448 (112 subjects in DMT, 185 subjects in NMT, and 151 subjects in FMT).

The model for those who found Sudoku difficult is significant ($F(8, 439)=24.93$; $p<.0001$). The results show that those in the forced condition performed significantly worse than those in either the discretionary or non-multitasking conditions (see Table 7-12). For a task that is considered more difficult and not easy, those in the forced multitasking condition performed significantly worse than those in the discretionary multitasking condition (H1a) and non-multitasking condition (H1b). If a task was hard for a subject, getting interrupted with additional tasks appeared to hurt performance. Therefore H1 was partially supported, since it was supported only for subjects who viewed Sudoku as more difficult.

Table 7-12. Comparisons of Multitasking Conditions where Sudoku Difficulty >= 3[12]

Multitasking Category	*Accuracy*
DMT	38.90 (A)
NMT	38.52 (A)
FMT	35.86 (B)
Model F(8, 439)	24.93***
R^2	31.24%
Explanatory Variables	
Condition	4.37*
Sudoku Experience	114.71***
Gender	15.59***
Age	22.65***
Computer Skills	0.68 ns
Academic Level	0.76 ns
Paid (vs. credit)	1.41 ns

Significance levels: * p<.05; ** p<.01; *** p<.001;
Duncan grouping (A or B) in parentheses.

7.3.2 Hypothesis 2

H2 predicted that subjects who did not multitask would perform significantly better in terms of productivity than forced multitaskers (H2a) and discretionary multitaskers (H2b). The productivity performance measures how much subjects were actually able to get done. When getting interrupted or interrupting oneself, people have to recall their thought process. Non-multitaskers, who had no interruptions, were therefore predicted to get the most accomplished. As Table 7-13 shows the model is significant (F(8, 626)=12.46;p<.0001) but not the means of the different conditions. Furthermore, the pattern tends to show that discretionary multitaskers,

[12] When performing the same statistical analysis where Sudoku difficulty is greater than 3 (and not greater than or equal to 3), this ANOVA is not significant. In addition, examining the data for only those subjects who considered the task to be easier (where Sudoku difficulty is less than or less than and equal to 3) is not significant.

rather than non-multitaskers, performed better in terms of productivity. Therefore, H2 is not supported.

Table 7-13. Comparison of Multitasking Conditions for Productivity

Multitasking Category	*Productivity*
DMT	60.24 (A)
NMT	58.31(A & B)
FMT	57.67 (B)
Model F(8, 626)	12.46***
R^2	13.73%
Explanatory Variables	
Condition	1.22 ns
Sudoku Experience	38.05***
Gender	6.39*
Age	20.23***
Computer Skills	0.02 ns
Academic Level	0.21 ns
Paid (vs. credit)	2.48 ns

Significance levels: * p<.05; ** p<.01; *** p<.001;
Duncan grouping (A or B) in parentheses.

In the table above, although the condition is not significant, some of the controls (such as Sudoku experience, gender and age) have a significant effect on productivity and make the overall productivity model significant.

The correlation between productivity and the level of difficulty of Sudoku is negative and significant for subjects in both the discretionary (ρ=-0.24962; p=0.0012) and forced (ρ=-0.18147; p=0.0081) conditions, but not the non-multitasking condition. This implies that in the forced and discretionary conditions when Sudoku was considered difficult, subjects performed worse in terms of their productivity. However, this was not the case with non-multitaskers. The productivity for subjects in the non-multitasking condition was not impacted by the level of difficulty of the Sudoku puzzle. When working in hard tasks, discretionary and forced multitasking subjects had a lower productivity, but non-multitasking subjects were not impacted by the level of difficulty.

In other words, if we compare the productivity for those who found Sudoku more difficult vs. those who found Sudoku easy it is significant that the former has a worse productivity than the latter for the forced multitasking condition ($F(1, 210)=7.49$, $p=.0067$) and discretionary multitasking condition ($F(1, 164)=13.32$, $p=.0004$), but the non-multitasking condition is not significant. When a task was considered hard, non-multitaskers did just as well in terms of productivity than when they considered the task easy (there were no significant differences). Perhaps during a more difficult task, receiving interruptions or deciding to multitask will have a negative impact on productivity. However, it is also possible that the condition affected people's responses on how they classified tasks that they considered harder or easier.

7.3.3 Hypothesis 3

H3 argued that when subjects were allowed to multitask at their discretion, those who switched more often would have a lower accuracy than those who switch less often. In this case, the number of switches was used as the independent variable and accuracy as the dependent variable. As shown in table 7-10, there were 166 subjects in the discretionary multitasking condition (as explained above 46 were moved into the non-multitasking condition). A correlation analysis was performed between number of switches and accuracy for the subjects in the discretionary condition and the result is negative and significant ($\rho=-0.24738$; $p=0.0013$).

This implies that as predicted the more subjects switched, the lower their scores in terms of accuracy.

Furthermore, a regression was performed with accuracy as the dependent variable against the number of switches and the other controls. The result shows that the model is significant (F(7, 158)=13.96; p<.0001) as well as the total number of switches which is negative and therefore depicted as a downward line (see Table 7-14). The more switches the lower the accuracy. Therefore, H3 is supported. When subjects had the option of switching and were in the DMT condition those that switched more often had lower scores. This finding is crucial to better understand the performance effects of multitasking and therefore this discretionary multitasking condition will be explored in more detail in section 7.4.

Table 7-14. Linear Model for Discretionary Condition Only

	Accuracy
Model F(7, 158)#	13.96***
R^2	38.21%
Parameters	*Estimate (t, p)*
Intercept	56.17 (7.63***)
TotalSwitches	-0.65 (-3.37**)
Sudoku Experience	4.24 (6.87***)
Gender	-5.03(-3.09**)
Age	-0.65 (-2.9**)
Computer Skills	-0.36 (-0.34 ns)
Academic Level	0.49 (0.46 ns)
Paid (vs. Credit)	0.62 (0.36 ns)

Significance levels: * p<.05; ** p<.01; *** p<.001;
Please note that in this test N=166

7.3.4 Hypothesis 4

Hypothesis 4 proposed that subjects who were in a forced multitasking environment and felt the primary task was difficult would have lower accuracy than those who felt the task was easy. In other words, getting interrupted during a task that one feels is difficult will hurt one's performance, while getting interrupted during a task that one feels is easy will not affect one's performance.

Consistent with the tests for H1, a more difficult task was considered as a task where the perceived Sudoku difficulty was rated as three or more. An easy task was a task where the level of difficulty for Sudoku was less than three. An Analysis of Variance was performed with accuracy as the dependent variable, in order to compare the effectiveness of those who considered the task easy versus those who considered the task to be harder in the forced multitasking condition. Table 7-15 shows that the model is significant ($F(7, 204)=18.14$, $p<.0001$) as well as the Sudoku difficulty variable. As the means show, in the forced condition, those who felt the task was harder performed significantly worse than those who felt the task was easier (35.86 vs. 44.45). This was only true for subjects in the forced multitasking condition. The difference in accuracy for subjects who considered the primary task easy or hard was not significant for the discretionary multitaskers or non-multitaskers (the model was significant but the Sudoku difficulty variable was not significant).

Table 7-15. Comparisons of Sudoku Level of Difficulty for the Forced Multitasking Condition

Multitasking Category	*Accuracy*
Easy	44.45 (A)
Hard	35.86 (B)
Model F(7, 204)	18.14***
R^2	38.36%
Explanatory Variables	
Sudoku Difficulty	19.42***
Sudoku Experience	57.27***
Gender	4.06*
Age	18.74***
Computer Skills	2.88 ns
Academic Level	1.03 ns
Paid (vs. credit)	0.86 ns

Significance levels: * p<.05; ** p<.01; *** p<.001;
Duncan grouping (A or B) in parentheses.

Although the influence of the perceived level of difficulty on accuracy might be viewed as obvious, it is worth noting that the accuracy measure also contains the scores in the secondary tasks. In addition to the total accuracy score, separate analyses were performed using only the accuracy for the primary task as well as the accuracy for only the secondary tasks. As expected in the forced multitasking condition, those who felt Sudoku was difficult had a lower accuracy than if they felt Sudoku was easier (see Table 7-16). Once again, the means for the level of difficulty were not significant in the discretionary and non-multitasking conditions.

Table 7-16. Comparisons of Sudoku Level of Difficulty for the Forced Multitasking Condition (Primary Task Only)

Multitasking Category	*Primary Tasks' Accuracy*
Easy	77.65 (A)
Hard	54.36 (B)
Model F(7, 204)	30.69***
R^2	51.30%
Explanatory Variables	
Sudoku Difficulty	22.58***
Sudoku Experience	144.50***
Gender	0.70 ns
Age	4.76*
Computer Skills	6.03*
Academic Level	5.37*
Paid (vs. credit)	4.08*

Significance levels: * p<.05; ** p<.01; *** p<.001;
Duncan grouping (A or B) in parentheses

While it can be viewed as obvious that those who found Sudoku harder performed worse on Sudoku, this was only true for forced multitaskers. Furthermore, when removing the primary task's accuracy from the accuracy variable and only examining the secondary tasks' accuracy the result is still significant (F(7, 204)=8.22, p<.0001, see Table 7-17). Therefore in the forced condition, subjects who considered Sudoku to be harder had a lower accuracy for the secondary tasks than those who considered Sudoku easier. This was not significant in the other two conditions. In the forced condition, when Sudoku appeared to be difficult it affected the subjects' performance on the secondary tasks.

Table 7-17. Comparisons of Sudoku Level of Difficulty for the Forced Multitasking Condition (Secondary Tasks Only)

Multitasking Category	*Secondary Tasks' Accuracy*
Easy	37.80 (A)
Hard	32.17 (B)
Model F(7, 204)	8.22***
R^2	21.99%
Explanatory Variables	
Sudoku Difficulty	9.04**
Sudoku Experience	12.15***
Gender	7.19**
Age	16.26***
Computer Skills	0.78 ns
Academic Level	0.03 ns
Paid (vs. credit)	0.04 ns

Significance levels: * p<.05; ** p<.01; *** p<.001;
Duncan grouping (A or B) in parentheses

In sum, receiving interruptions when the primary task was difficult did not only affect the primary task, but the secondary tasks as well. Accuracy was significantly lower when forced multitaskers considered Sudoku harder than when they considered it easier when examining all three of the following cases: Sudoku only, the secondary tasks only, and the combination of all of the tasks. Therefore H4 is supported.

7.3.5 Hypothesis 5

H5 predicted that the more one considered a task as difficult, the more switches s/he would have. However, when correlating the Sudoku difficulty variable from the post-test questionnaire with the number of switches in the discretionary multitasking condition, it is not significant. But when correlating the Sudoku difficulty with *Sudoku switches* it is significant but negative (ρ=-0.18912, p=0.0147). Sudoku switches are different than total switches in that rather than calculating the total number of tab clicks, this variable only calculates the total number of Sudoku tab clicks. An additional variable which calculates the amount of time the Sudoku task was switched *from* (instead of switched *to*) was also significant and negative (ρ=-0.20282, p=0.0088). Those subjects who felt that the Sudoku task was difficult actually had fewer Sudoku switches both to and from Sudoku, and if they found the task easier they had more Sudoku switches. This implies that they tended to stay on the Sudoku task if they found it difficult. This is the exact opposite of the hypothesis. Rather than having more Sudoku switches when subjects felt that the task was hard, they tended to stay on the task.

7.3.6 Hypothesis 6

H6 predicted that forced multitaskers would have more frustration than non-multitaskers (H6a) and discretionary multitaskers (H6b). While the model is significant (F(8, 626)=4.25, p<.0001), the condition is not significantly different as shown in Table 7-18.

Table 7-18. Comparison of Multitasking Conditions for Frustration

Multitasking Category	*Frustration*
NMT	2.65 (A)
FMT	2.65 (A)
DMT	2.62 (A)
Model F(8, 626)	4.25***
R^2	5.15%
Explanatory Variables	
Condition	0.12 ns
Sudoku Experience	8.98***
Gender	1.00 ns
Age	0.02 ns
Computer Skills	18.54***
Academic Level	0.18 ns
Paid (vs. credit)	0.47 ns

Significance levels: * p<.05; ** p<.01; *** p<.001;
Duncan grouping (A or B) in parentheses.

The table also shows that two of the control variables, Sudoku experience and computer skills, do affect a subject's level of frustration. When correlating Sudoku experience with frustration, in the non-multitasking condition, the more experience one had the less frustration they experienced (ρ=-0.24669, p<.0001). In fact, when examining only those users who never played Sudoku before (Sudoku experience = 0), then the forced multitaskers actually have significantly less frustration than the non-multitaskers ($Mean_{NMT}$=2.86 vs. $Mean_{FMT}$=2.63; F(1,181)=5.43, p=.0209). When correlating frustration with computer skills it is negative and significant for each condition. Those with less computer skills experienced more frustration (FMT: ρ=-0.21476, p=0.0017; DMT: ρ=-0.19604, p=0.0114; NMT: ρ=-0.15399, p=0.0133).

In terms of Sudoku difficulty, when examining only subjects who felt Sudoku was easier or subjects who felt Sudoku was more difficult, the frustration variable is still insignificant. There is also no significant correlation between frustration and Sudoku difficulty.

Furthermore, the frustration variable is significant and negatively correlated with accuracy in both the discretionary (ρ=-0.15838; p=0.0416) and non-multitasking (ρ=-0.13710; p=0.0277) conditions, but not with the forced condition. In other words, subjects with lower accuracy had more frustration in the discretionary and non-multitasking conditions, whether this is due to the fact that they had lower scores which caused the increase in frustration or were frustrated and therefore performed worse. However, in the forced multitasking condition frustration did not correlate with accuracy. Therefore, H6 is not supported.

7.4 Additional Analyses

7.4.1 Discretionary Multitasking Condition Sub-Sample

The discretionary multitasking condition, more so than any other condition, was examined in greater depth for two reasons. First, previous research has not studied in depth the performance effects of people who voluntarily choose to multitask, and second, this condition had the most flexibility for analyzing the performance effects of varying degrees of multitasking. Since the number of switches in this condition varied depending on the subject's desire to multitask, the performance of those who barely multitasked versus those who multitasked much more can be compared.

There were 212 subjects in the discretionary multitasking condition (including those subjects who decided not to multitask). About half (103) were paid to do the experiment ($10 for their participation), while the remaining participants (109) received course credit. Those who worked for monetary payment had a lot more switches than those who were compensated with class credit ($Mean_{Paid} = 9.43$ vs. $Mean_{Credit} = 7.23$, $F=15.17$, $p=0.0001$). Furthermore, subjects who participated for payment reported a higher tendency to multitask than those in the credit sub-sample ($mt\_propensity_{Paid}=3.83$ vs. $mt\_propensity_{Credit}=3.56$, $F=7.70$, $p=0.006$).

As explained previously, the minimum amount of switches that any subject could have was five, which indicated that the subject chose not to multitask and to perform the six tasks in a sequential manner. While the number of switches for the paid group ranged from 5-29, the number of switches for the credit group only ranged from 5-20. Comparatively speaking, most of those in the credit group are found in the lower end of the switch range, while those in the paid group are more spread out (60% had only 5 and 6 switches compared to the 38% in the paid group). Therefore, an analysis of only the paid group in the discretionary multitasking condition was examined in order to compare the differences in performance between those who had a high number of switches versus those who had a lower number of switches. See Table 7-19 for a breakdown of subjects across the paid vs. credit groups.

Table 7-19. Breakdown of Discretionary Subjects in Paid vs. Credit Groups

	Discretionary Multitasking Condition
Paid	103*
Credit	109

*In this discretionary sub-analysis section, we will only be examining the 103 paid subjects.

Descriptive statistics of the variables of interest for participants in the discretionary paid condition are presented in Table 7-20.

Table 7-20. Means of Strategy and Performance Variables in Discretionary Condition Only

[N=103]	*Mean*	*Std Dev*	*Minimum*	*Maximum*
Number of Switches	9.43	5.09	5	29
Accuracy	41.11	13.14	12.94	69.66
Productivity	61.14	12.90	31.89	93.33

Different multitasking strategies, discretionarily applied by participants in the discretionary condition, were captured by the number of switches and illustrated with timeline graphs. Sixteen subjects did not multitask and applied a strictly sequential working strategy. (Note: there were 46 non-multitaskers in the discretionary sample, but only 16 were from the paid group. The remainder was from the credit group. These 16 users were left in this discretionary analysis.) Figures 7-1 and 7-2 show examples of two distinct working patterns. Figure 7-1 shows the timeline graph of a subject who chose not to multitask and allocated his/her time in a completely sequential fashion, with only five switches. In the timeline graph, the line segments indicate the amount of time (in seconds) that each task was active based. In this case, the subject first completed the Sudoku task, followed by the Word task, then the Number Series task, the Visual task, the other Number Series task, and finally the other Visual task.

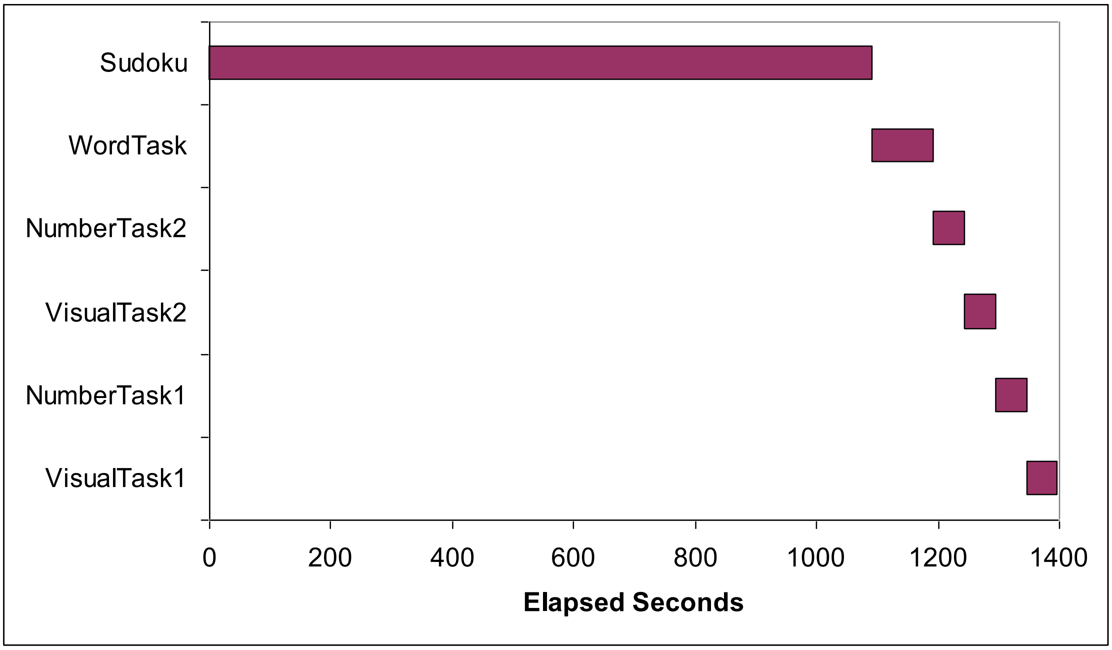


Figure 7-1. Participant Applying a Sequential Strategy (Not Multitasking)

Figure 7-2 shows the timeline of a subject who had 10 switches and therefore had more multitasking than the non-multitasker. In this case, after about two minutes playing Sudoku, the subject switched to the first visual task, completed it and went on to start and finish the word task and the number series task. The subject then began the other visual task, but before completing it switched to the first number series task which s/he completed. An examination shows that this participant then went back and forth between playing Sudoku for a while and switching to the visual task. When time ran out in the visual task, s/he spent the last minutes working on the Sudoku problem. The application controls the total time on a tab according to the time limits. If the time allotted for a task has not run out, the task is still active and the participant can return to the tab as many times as s/he wishes.

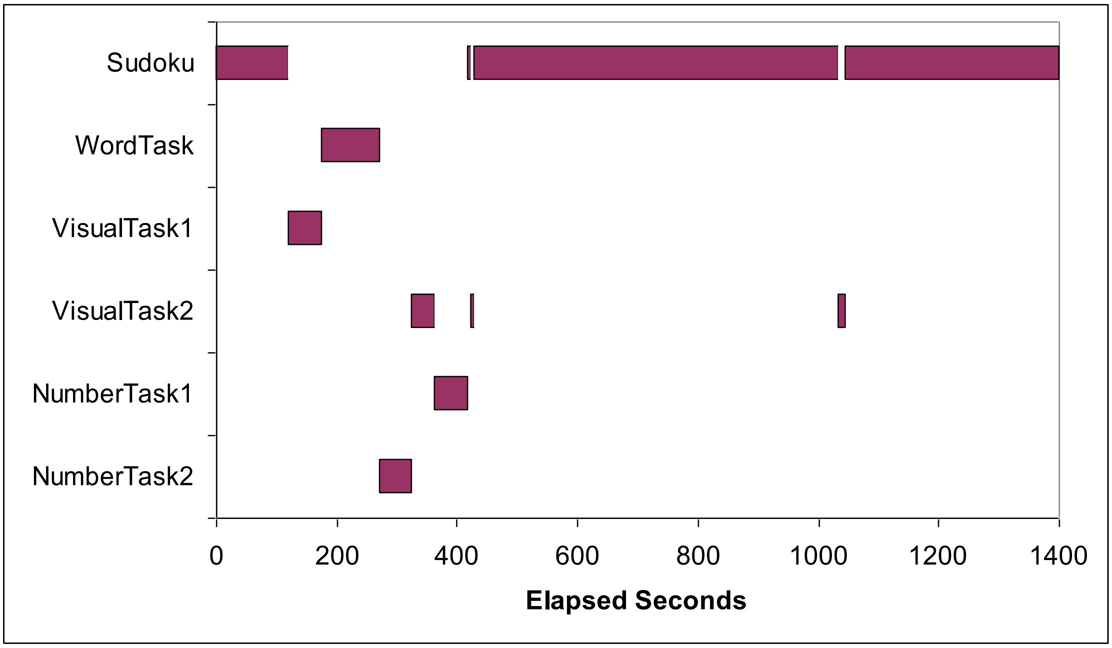


Figure 7-2. Participant Applying an Interleaving Strategy

Recall that H3 was supported in the large discretionary sample. Those with more switches performed worse in terms of accuracy. To examine this again for only the paid sample and in greater detail, a linear model was run on the discretionary subsample examining accuracy and number of switches. In the large discretionary multitasking sample, productivity is not significantly correlated with number of switches. Productivity, in addition to accuracy was examined for the discretionary paid sample as a dependent variable. With both models, the number of switches is the main independent variable. The models also included the control variables (age, gender, computer skills, academic level, and Sudoku experience).

Table 7-21 presents the results of the model.

Table 7-21. Linear Model

	Accuracy	*Productivity*
Model F(5, 97)[#]	12.62***	1.40 ns
R^2	44.09%	8.02%
Parameters	*Estimate (t, p)*	
Intercept	29.15 (2.65**)	58.48 (4.23***)
TotalSwitches	-0.41 (-2.05*)	-0.05 (-0.18 ns)
Sudoku Experience	5.27 (7.20***)	2.21 (2.40*)
Gender	-5.67(-2.77**)	-1.41 (-0.55 ns)
Age	-0.37 (0.80 ns)	0.22 (0.37 ns)
Computer Skills	1.81 (1.39 ns)	0.28 (0.17 ns)
Academic Level	-1.73 (-1.19 ns)	-1.93 (-1.06 ns)

Significance levels: * p<.05; ** p<.01; *** p<.001;
The N for this model is 103

For number of switches (the main dependent variable), only the linear model for accuracy was significant with p<.001. Figure 7-3 shows a simplified graph of the linear model for accuracy, where X is the number of switches and Y is accuracy. For visualization and ease of interpretation purposes, the plotted curve corresponds to Y = 29.15 – 0.41X and the control variables are omitted for simplicity. Therefore, H3 was supported in this paid sample as well. As multitasking increased, accuracy decreased.

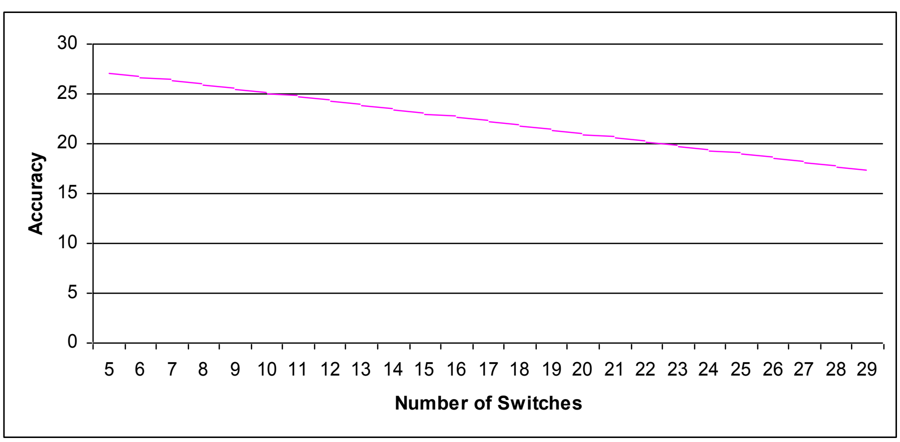


Figure 7-3. Graph of Linear Model for Accuracy for Discretionary Paid Sample

In terms of productivity, an inverse relationship was not significant. However, when testing a quadratic model instead of a linear model, the results are significant.[13] Table 7-22 presents the results of the model in terms of productivity as well as the R^2, the estimated coefficients, and the significance of each one.

Table 7-22. Quadratic Model

	Productivity
Model F(6, 96)#	2.22*
R^2	14.08%
Parameters	*Estimate (t, p)*
Intercept	43.85 (3.01***)
TotalSwitches	2.45 (2.46*)
TotalSwitchesSquared	-0.09 (-2.59*)
Sudoku Experience	2.00 (2.22*)
Gender	-2.26(-0.89 ns)
Age	0.28 (0.50 ns)
Computer Skills	0.68 (0.42 ns)
Academic Level	-2.24 (-1.26 ns)

Significance levels: * p<.05; ** p<.01; *** p<.001;
The N for this model is 103

Both the quadratic term and the linear term are significant (coefficients -0.09, p<.05 and 2.45, p<.05 respectively). Furthermore, the sign of the coefficient of the quadratic term is negative indicating an inverted-U relation.

Figure 7-4 shows a simplified graph of the quadratic model for productivity. Once again, for visualization and ease of interpretation purposes, the plotted curve corresponds to $Y = 43.85 - 0.09X^2 + 2.45X$ and the control variables are omitted for simplicity. The inverted-U pattern is clearly observed in this graph. What this signifies is that those who multitasked did significantly better in terms of productivity than those who did not or only slightly multitasked, but up to a

[13] In terms of accuracy, a quadratic model was not significant. As stated previously though there was an inverse relationship between accuracy and the number of switches.

point. At some point there were diminishing returns where too much multitasking hurt productivity.

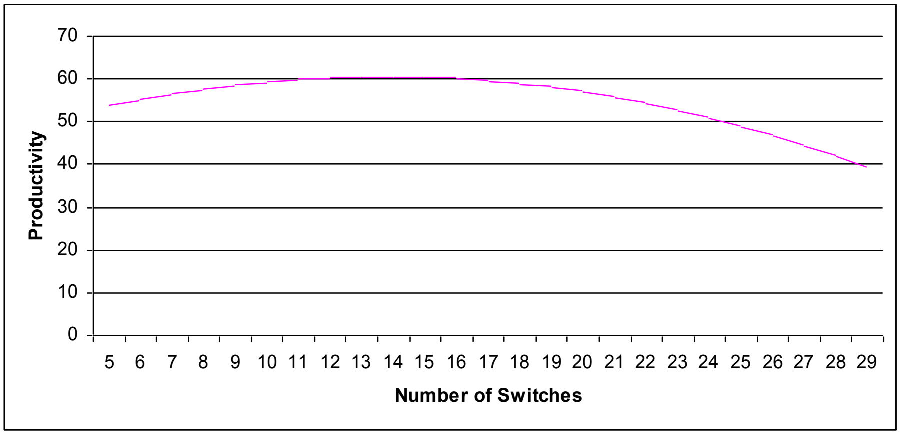


Figure 7-4. Graph of Quadratic Model for Productivity for Discretionary Paid Sample

For a more detailed analysis of performance within the discretionary/paid sub-sample, the number of switches was divided into three different categories: Low (number of switches below 10), Medium (number of switches between 10 and 15) and High (15 or more switches). The cutoff points (10 and 15) were determined from the mean and standard deviation of this variable for the discretionary sub-sample (shown in Table 7-20). These three categories represent the extent to which participants engaged in multitasking behavior, whether they were light multitaskers, medium multitaskers, or heavy multitaskers.

With an analysis of variance, the differences in accuracy and productivity for these three categories of multitaskers were examined. As shown in Table 7-23, the two ANOVA models are significant. All models were run with age, gender, computer skills, academic level, and Sudoku experience as controls.

Table 7-23. Analysis of Variance Results (Discretionary Sub-Sample)

Multitasking Category	Accuracy	Productivity
Low	41.62 (A)	59.17 (B)
Medium	47.05 (A)	71.20 (A)
High	30.89 (B)	56.66 (B)
Model F(7, 95)	11.53***	3.62**
R^2	45.93%	21.06%
Explanatory Variables		
Multitasking Category (L,M,H)	3.76*	7.86***
Sudoku Experience	41.83***	2.89 ns
Gender	7.40**	0.36 ns
Age	0.69 ns	0.22 ns
Computer Skills	2.70 ns	0.50 ns
Academic Level	1.53 ns	1.88 ns

Significance levels: * p<.05; ** p<.01; *** p<.001;
Duncan grouping (A or B) in parentheses.

In the ANOVA model of accuracy, a Duncan analysis showed two significantly different groups. Those in low and medium multitasking levels had comparable levels of accuracy, while those at high multitasking levels experienced a significant decline in the percentage of correct answers. See Figure 7-5. While the figure appears to be an inverted U, in effect it is a downward line, since the mean of the low (41.62) and medium (47.05) multitaskers are statistically similar (i.e. no significant differences between them as shown by the same Duncan grouping A), while the mean of the heavy multitaskers (30.89) is significantly lower than both the low and medium multitaskers.

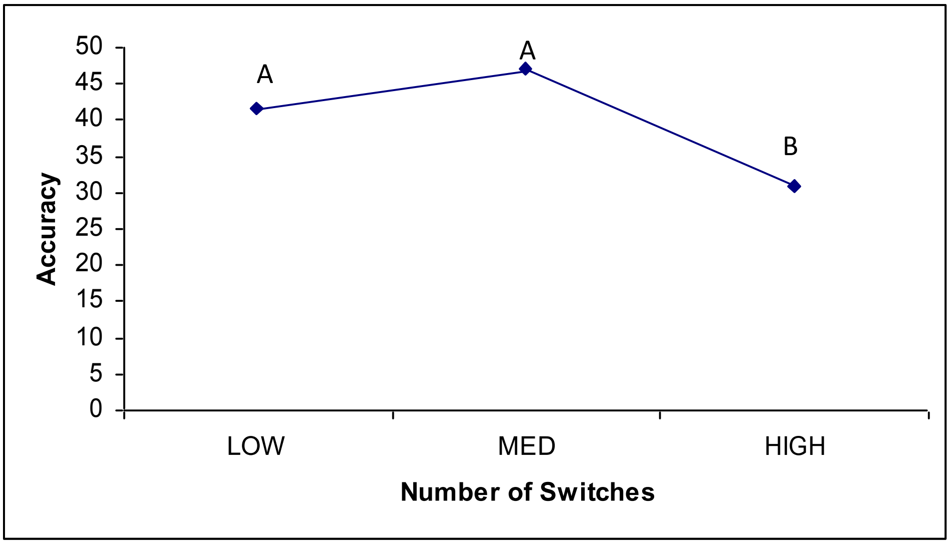


Figure 7-5. Accuracy and Multitasking Activity for Discretionary Paid Sample

In the ANOVA model of productivity, a medium multitasking level is associated with the highest productivity, while the extremes (low and high) correspond to lower productivity. The results for productivity are consistent with the prediction of the inverted-U curve, as they show two significantly different groups. The most productive are those with moderate levels of multitasking, while those at the extremes with low or high multitasking are least productive. Figure 7-6 shows a graph of the productivity score averages in each category. This graph produces an inverted-U pattern as predicted by our hypothesis. Low multitaskers had an average productivity score of 59.17 while medium multitaskers' productivity score was significantly better at 71.20. In contrast, heavy multitaskers exhibit significantly worse productivity scores (56.66).

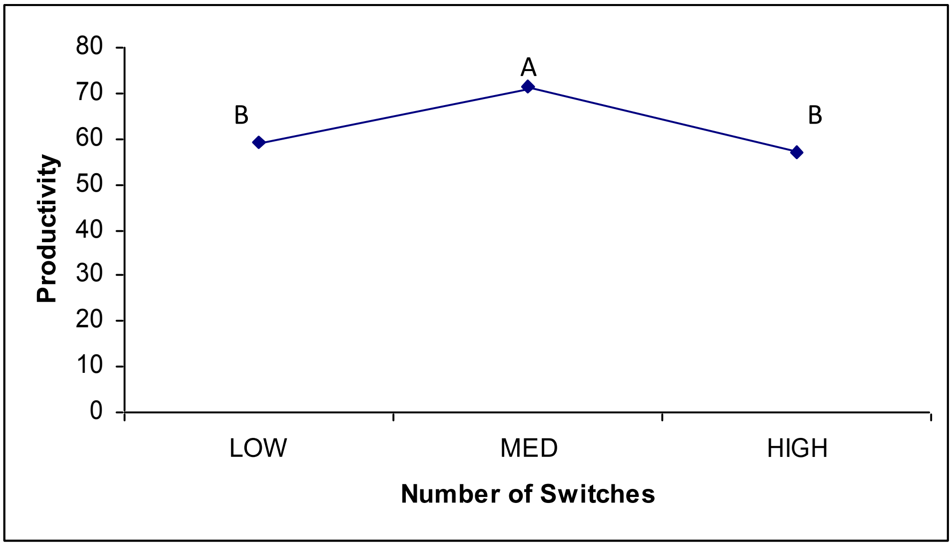


Figure 7-6. Productivity and Multitasking Activity for Discretionary Paid Sample

Productivity is the measure of the completion for each task. While people may feel very productive since they are accomplishing more when multitasking, it is an illusion. Multitasking may help in completing tasks, however in terms of accuracy it is all downhill.

7.4.2 Non-Multitasking Condition Sub-Sample

The discretionary group has shown us the effect performance has when people choose to multitask. This DMT condition had the most flexibility since the number of switches was variable and subjects were able to decide how many times to switch to another task. In the non-multitasking condition the number of switches was always 5 for all subjects and in the forced multitasking condition the number of switches was always 10. While there was not as much that could be analyzed alone within these two conditions, however there were some surprising observations, especially within the non-multitasking condition.

The propensity to multitask variable indicates the extent to which one is inclined to multitask in general. This variable was acquired from the pre-test questionnaire, which asked subjects whether they tended to multitask in general. It is important to analyze the congruence between the subjects' preferences to multitask and the actual multitasking condition to which they were assigned. Prior research has shown that subjects tend to perform better when there is congruence between their preferences and behavior (Lee 1999; Slocombe and Bluedorn 1999).

Regardless of their propensity to multitask, subjects were randomly divided into the three conditions. Subjects in the non-multitasking condition were not allowed to multitask. Non-multitaskers who had a propensity to multitask in general performed better than subjects who did not have a propensity to multitask: multitasking propensity was positively correlated with accuracy ($\rho=0.17535$, $p=0.0047$) as well as productivity ($\rho=0.19382$, $p=0.0018$). When a subject tended to multitask in general, if they were put in the non-multitasking condition (and therefore did not multitask) their performance was better than those who tended not to multitask in general. Therefore, the non-multitasking condition was divided into two groups, those who have a propensity to multitask in general and those who do not multitask in general.

There were 46 discretionary multitaskers who were moved into the non-multitasking condition. This decision was made in order to compare the performance between multitaskers and non-multitaskers. However, it should be noted that these 46 subjects are different psychologically than the non-multitaskers in that they were allowed to make a decision on whether they could or could not multitask. Therefore the above analysis was performed both with and without the 46 discretionary non-multitasking users in the non-multitasking sample and it was significant in both ways.

In order to perform additional analyses of accuracy and productivity using this distinction, analyses of variance were performed to determine the point at which those with a propensity to multitask had a better accuracy and productivity than those with less of a propensity to multitask.

Using the multitasking propensity mean (3.5) it was significant that those with a multitasking propensity of greater than 3.5 (high tendency to multitask) had better performance than those with less of a multitasking propensity. Therefore, this value was used as the cut-off.

The term *I-NMT* (Incongruent Non-Multitaskers) will be used to describe subjects who like to multitask but could not (i.e. those with a propensity to multitask but cannot). *C-NMT* (Congruent Non-Multitaskers) will be used to describe the remaining non-multitaskers who prefer to perform one task at a time.

7.4.3 Forced Multitasking Condition Sub-Sample

We saw in the previous section that in the non-multitasking condition, subjects with a greater propensity to multitask performed better both in terms of accuracy and productivity than subjects who had a lower tendency to multitask. In the forced multitasking condition, subjects who tended to multitask in general did better than those who did not have a propensity to multitask but only in terms of productivity. Multitasking propensity was positively correlated with productivity ($\rho=0.14948$, $p=0.0296$). This was not significant in terms of accuracy. In the discretionary sample, neither dependent variable was significantly correlated with multitasking propensity.

In order to separate those subjects out those who do not like to multitask and therefore performed worse in terms of productivity, analyses of variance were conducted to determine the point at which those with a propensity to multitask performed better than those who did not multitask in general. Using the mean (as in the case with the non-multitasking condition) did not produce significant results. However, based on the mean and standard deviation the value 3 was chosen and this was significant that those with greater than three multitasking propensity had a better productivity than others in that condition.

Therefore, the term *I-FMT* (Incongruent Forced Multitaskers) will be used to describe subjects who do not like to multitask (low multitasking propensity) but were forced to multitask in this condition, while *C-FMT* (Congruent Forced Multitaskers) will be used to describe the remaining forced multitaskers who prefer to multitask. Note: this category will only be used when describing productivity since there was no correlation between accuracy and multitasking propensity.

Different monochronic or polychronic behaviors can be associated with both individuals' preferences and with the events and tasks themselves; this can be described in terms of a 2 X 2 matrix (Lee 1999). See Figure 7-7 for a description of individuals' propensity to multitask and the condition they were assigned to.

		Condition	
MT Propensity		**NMT**	**FMT**
	Low	C-NMT	I-FMT
	High	I-NMT	C-FMT

Figure 7-7. Four Possibilities of Individuals' Propensities and Conditions

7.4.4 Analysis of Accuracy with All Three Multitasking Conditions

Since non-multitaskers are very divided in terms of those with a propensity to multitask who are not able to multitask (I-NMT) and those who are naturally not inclined to multitask anyway (C-NMT), an analysis of variance was conducted comparing the accuracy of all three conditions with this distinction for non-multitaskers only. Table 7-24 shows the results. The model is significant (F(9, 625)=38.97; p<.0001) as well as the condition (F=3.36; p=0.02).

Table 7-24. Further Comparisons of Multitasking Conditions

Multitasking Category	*Accuracy*
I-NMT	42.31 (A)
DMT	40.53 (A & B)
FMT	38.33 (B)
C-NMT	35.72 (C)
Model F(9, 625)	38.97***
R^2	35.95%
Explanatory Variables	
Condition	3.36*
Sudoku Experience	215.29***
Gender	27.56***
Age	24.66***
Computer Skills	5.12*
Academic Level	0.11 ns
Paid (vs. credit)	0.32 ns

The above table shows that subjects with a high multitasking propensity who were assigned to the NMT condition had the highest accuracy, while subjects with a low propensity in that condition had the lowest accuracy. The second lowest group was those subjects who were forced to multitask in the FMT condition. However, when examining the results of the discretionary DMT condition it is divided between Duncan groups A & B.

In the discretionary condition sub-analysis section, we saw that there were performance differences within the discretionary condition (in the paid sub-sample). Depending on the multitasking category (low, medium or high) the performance was different. Therefore, it is possible that this average is masking the existence of significant differences. To examine this possibility, high multitaskers (who performed worse than both low and medium multitaskers in terms of accuracy in the discretionary analysis) were analyzed separately when examining accuracy across conditions.

The term *H-DMT* will be used to describe high multitaskers, i.e. those with 15 or more switches (this value was chosen from the discretionary analysis section although now it includes the larger sample), while *DMT* is the remaining discretionary multitaskers. Table 7-25 shows an analysis of variance for the accuracy scores for the new categories. The model is significant ($F(10, 624)=36.91$; $p<.0001$) as well as the condition ($F=5.59$; $p=.0002$). Those who did not multitask, but had a propensity to multitask in general and those in the discretionary condition who were able to choose how to multitask performed significantly better than all the rest. Those who were forced to multitask and the remainder of the non-multitaskers (those who do not generally multitask too much and were in the non-multitasking condition) did significantly worse than the above two conditions. While forced multitaskers had lower accuracy than discretionary multitaskers, those in the high multitasking group, who had a lot of switches, had the lowest accuracy from everyone. See Figure 7-8 for a graph representation of Table 7-25.

Table 7-25. Further Comparisons of Multitasking Conditions

Multitasking Category	*Accuracy*
I-NMT	42.31 (A)
DMT	41.92 (A)
FMT	38.33 (B)
C-NMT	35.72 (B)
H-DMT	29.83 (C)
Model F(10, 624)	36.91***
R^2	37.16%
Explanatory Variables	
Condition	5.59***
Sudoku Experience	203.91***
Gender	28.09***
Age	25.09***
Computer Skills	5.96*
Academic Level	0.31 ns
Paid (vs. credit)	0.98 ns

Significance levels: * p<.05; ** p<.01; *** p<.001;
Duncan grouping (A or B) in parentheses.

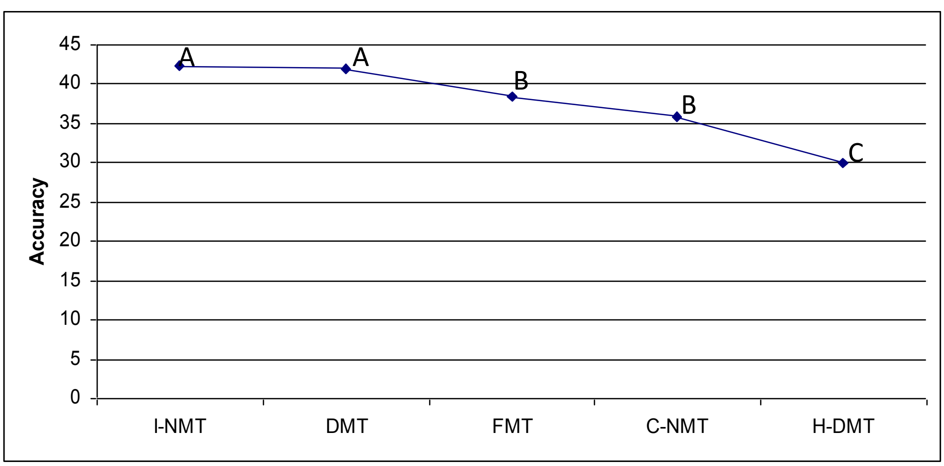


Figure 7-8. Accuracy Across Conditions

7.4.4 Analysis of Productivity with All Three Multitasking Conditions

In terms of productivity, as discussed previously those subjects who were forced to not multitask (i.e. they have propensity to multitask in general) were separated out from the natural non-multitaskers who do not multitask in general. In addition, in the forced multitasking condition, those who do not like to multitask (low propensity to multitask) and were forced to multitask (I-FMT) were separated out from those who like to multitask and were forced to multitask (C-FMT).

When discussing accuracy high discretionary multitaskers were separated out, but we saw in the discretionary analysis section that regarding productivity, middle multitaskers performed better than both high and low multitaskers. Therefore, middle multitaskers are now analyzed separately. The term M-DMT will be used to describe middle discretionary multitaskers.

See Table 7-26[14] for an analysis of variance for the productivity scores with the new categories. The most noteworthy observation is that middle multitaskers had the best productivity (better than those who were forced to multitask, or who did not multitask, or who had a low or high amount of multitasking, etc.). See Figure 7-9 for the graphical representation.

[14] As mentioned previously, the DMT condition could not be categorized in terms of multitasking propensity for either accuracy or productivity since there were no significant correlations between multitasking propensity and performance for subjects in this condition.

Table 7-26. Further Comparisons of Multitasking Conditions

Multitasking Category	*Productivity*
M-DMT	64.97 (A)
I-NMT	60.80 (B)
C-FMT	59.11 (B)
DMT	59.11 (B)
C-NMT	54.59 (C)
I-FMT	54.17 (C)
Model F(11, 623)	10.45***
R^2	15.57%
Explanatory Variables	
Condition	3.20**
Sudoku Experience	32.53***
Gender	6.54*
Age	18.29***
Computer Skills	0.42 ns
Academic Level	0.55 ns
Paid (vs. credit)	1.78 ns

Significance levels: * p<.05; ** p<.01; *** p<.001;
Duncan grouping (A or B) in parentheses.

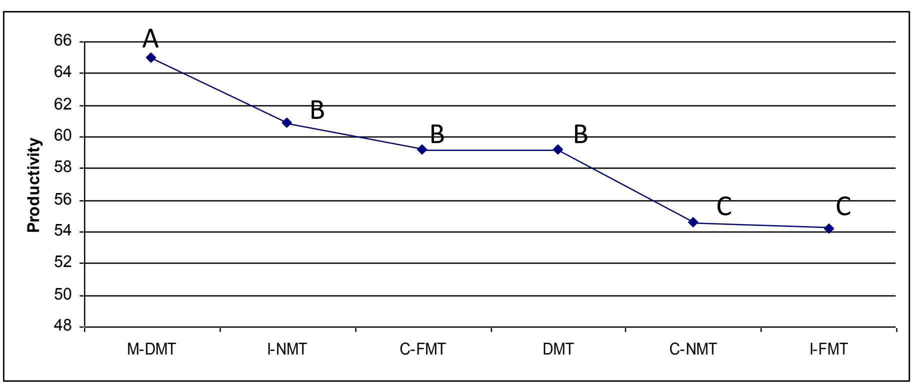


Figure 7-9. Productivity Across Conditions

7.4.5 Why People multitask?

In terms of accuracy, the results indicate that heavy multitasking caused performance to decrease. Why then do people multitask? Subjects in the discretionary condition had an open-ended question asking why they multitasked (Appendix I). This was done to help determine why people multitask and whether their performance varied depending on their reason for switching.

In the entire sample, there were 212 discretionary multitasking subjects. Analysis of the open-ended post-test questionnaire responses indicated that 26 subjects were not aware that they were free to switch tabs at will.[15] Data from these subjects was removed from the sample. The qualitative data to the open-ended question of "why did you switch?" or "why didn't you switch?" for the remaining 186 participants was analyzed. Two independent coders classified the answers into categories. The inter-coder reliability calculated with the percentage of agreement was 91%. The discrepancies between the coders were solved by discussing the differences and achieving agreement in the classification. After the coding was finalized, twenty answers were coded as 'Other' and subjects with this code were also removed from the sample (see Table 7-27).

Table 7-27. Types of Self-Interruptions

Why Switched	*Discretionary Subjects*
Did Not Know They Could Switch	26
Provided a Clear Reason that was Categorized	166*
Categorized as "Other"	20
Total Discretionary Subjects	**212**

*166 subjects were able to be classified into categories in order to analyze why people chose to multitask.

[15] While 26 subjects wrote that they did not know they could multitask this was not the number used to determine whether those subjects should be moved into the non-multitasking condition. Since many of the 26 users actually had above 5 switches, their actual number of switches was used to ascertain those who actually multitasked from those who did not.

Table 7-28 shows the different reasons that subjects multitasked and how many subjects were in each category. For the rest of this section, N is the 166 subjects that were categorized into groups.

Table 7-28. Types and Reasons for Self-Interruptions

	Freq (%)	*Descriptions*	*Examples of illustrative quotes from participants*
Negative Reasons			
Obstruction	30 (18%)	Temporary roadblock in the performance of a task.	*When I couldn't figure out what I was doing wrong*
Exhaustion	14 (8%)	Person experiences cognitive fatigue.	*When I needed to clear my head, I would switch to another task*
Frustration	10 (6%)	Ongoing task is too difficult given the level of skills.	*I switched because the Sudoku was hard*
Positive Reasons			
Reorganization	31 (19%)	Restructuring of the workload to improve performance.	*To finsih* [sic] *the easiest one first*
Exploration	8 (5%)	Opportunistic behavior to discover enticing alternative tasks.	*Just to see the actual content of each task*
Boredom	6 (4%)	Ongoing task is too easy given the level of skills.	*When I get bored from the task and wasn't giving proper attention*
No self-interruptions			
Focus-Task-Strategy	67 (40%)	Performed tasks sequentially	*I was focused on completeing* [sic] *one task at a time*
Total	***166 (100%)***		

The categories for switching were divided into switching for negative reasons or positive reasons. Negative reasons for switching included frustration, obstruction, and exhaustion. *Frustration* consisted of people switching due to the task being too difficult and/or causing them anxiety or frustration. This caused them to switch to another task. *Obstruction* occurred when subjects were stuck and could not figure out what was wrong so they chose to switch to another task. *Exhaustion* happened when subjects were tired and needed to clear their head for a bit and then return to the previous task. All these three categories were grouped under *negative* reasons for switching. Negative because the subjects could not complete the task and had to switch tasks because of that fact.

Positive reasons included boredom, reorganization, and exploration. *Boredom* included those subjects who were bored, such as a subject who is over-qualified for a task and found it too easy and was therefore bored. *Exploration* consisted of switches out of curiosity. There were subjects who wrote that they switched just to see all of the tasks first. *Reorganization* included subjects who switched in order to organize the content in a way to make them achieve better results, such as getting the smaller tasks out of the way or doing the tasks that they felt were easier first. These three categories were considered *positive* reasons for switching, since they generally consisted of subjects switching not because of tasks being too difficult, but rather these were subjects whose responses seemed more in control of the tasks and its level of difficulty and were able to proceed accordingly.

Overall, as shown above in Table 7-28, one third of the sample switched tasks for negative reasons, 27% did so for positive reasons, and the rest of the participants (40%) indicated that they avoided self-interruptions due to their deliberate strategy to focus on one task at a time.

The descriptive statistics of the quantitative variables for these 166 discretionary subjects are shown in Table 7-29.

Table 7-29. Descriptive Statistics

[n=166]	*Mean*	*S.D.*	*Min*	*Max*
Sudoku Skills	1.61	1.43	0	5
Number of Switches	8.32	4.12	5	28
Accuracy	40.67	13.13	12.94	69.66
Productivity	59.39	13.78	20.20	93.33

To test whether the stated qualitative reasons were indeed reflective of different patterns of activity during the experiment, the number of switches was checked. For this validation, the sample was divided into two groups, those who indicated that they limited their switches vis-à-vis those who mentioned any other reason for their switching (positive or negative). The average number of switches between these two groups is significantly different according to the results of a t-test (No-Self-Interruption Mean = 6.46 vs. Other-reasons Mean = 9.58; $t = 5.14$; $p<.0001$). This comparison provides an initial validation of the qualitatively collected self-interruption data with respect to actual patterns of behavior.

An ANOVA on the number of self-interruptions by type of trigger was conducted to test whether the participants' amount of self-interruptions varied depending on the nature of the trigger. The model is significant ($F(2, 163)=15.35$; $p<.0001$). Results shown in Table 7-30 indicate that those who experienced negative feelings switched more often than those in the positive category.

Table 7-30. Number and Type of Self-Interruptions

Type of Self-Interruption Trigger	*Number of Self-Interruptions*
Negative	10.26 (A)
Positive	8.76 (B)
No-Self-Interruptions	6.46 (C)
Model F(2, 163)	15.35***
R^2	16%

Significance level ***p<.001
Duncan group (A, B or C) in parentheses.

For the overall performance analysis, an ANOVA model of the participants' reasons for switching was used. The analysis was focused on those who switched (i.e. it excluded participants who indicated no self-interruptions). As shown in Table 7-31, the model is significant in terms of accuracy (F(7, 91)=8.43; p<.0001). A Duncan analysis produces two distinct groups. The highest overall accuracy among the multitaskers was obtained by those in the positive trigger group, while the lowest score is for those in the negative trigger group. This shows that those who switched for negative reasons had a lower accuracy than those who switched for positive reasons (productivity was not significant).

Table 7-31. Performance Analysis

Type of trigger	*Accuracy*	*Productivity*
Positive	42.98 (A)	62.38 (A)
Negative	37.16 (B)	57.74 (A)
Model F(7, 91)	8.43***	2.05 ns
R^2	39.32%	13.64%
Type of Trigger	4.75*	3.08 ns
Sudoku Experience	38.24***	4.23*
Gender	8.38**	1.38 ns
Age	0.06 ns	1.05 ns
Computer Skills	1.09 ns	0.72 ns
Academic Level	0.22 ns	0.04 ns
Paid	0.65 ns	0.00 ns

Significance levels *p<.05 *p<.01 ***p<.001;
Duncan group (A or B) in parentheses.

7.5 Summary

To summarize, two hypotheses were supported (H3 and H4), one was partially supported (H1) and three were not supported (H2, H5, and H6), as shown in Table 7-32. Regarding H1, examining accuracy across conditions showed no differences in performance. However, when removing those who thought Sudoku was easy, the accuracy in the forced multitasking condition was significantly lower than both the discretionary multitasking condition (H1a) and the non-multitasking condition (H1b). Therefore, H1, H1a, and H1b are partially supported.

Regarding H2, those in the non-multitasking condition did not appear to complete more of the tasks than those in the forced multitasking condition (H2a) or the discretionary multitasking condition (H2b). Thus, H2 was not supported.

In contrast, both H3 and H4 were supported. Consistent with H3, subjects in the discretionary multitasking condition that had more switches had lower accuracy than those who had less switches. H4 was also supported. Those in the forced multitasking condition who felt the task was more difficult had lower accuracy than those who did not feel the task was difficult. This was not the case for the discretionary and non-multitasking conditions.

Another hypothesis that was not supported was H5. The results were significant but in the opposite direction. When a subject in the discretionary multitasking condition considered a task as more difficult, s/he tended to stay on that task and switch less often.

Finally, H6 was not supported. Subjects in the forced multitasking condition did not experience more frustration than those in either the non-multitasking (H6a) or discretionary multitasking (H6b) condition.

Table 7-32. Summary of Hypotheses Results

H #	*Hypothesis*	*Results*
H1	Mandatory interruptions will have a negative effect on accuracy.	*Partially Supported*
H1a	Those who multitask at their discretion will have greater accuracy than those forced to multitask.	*Partially Supported*
H1b	Those who do not multitask will have greater accuracy than those forced to multitask.	*Partially Supported*
H2	Those who do not multitask will have greater productivity than those who multitask.	*Not Supported*
H2a	Those who do not multitask will have greater productivity than those forced to multitask.	*Not Supported*
H2b	Those who do not multitask will have greater productivity than those who multitask at their discretion.	*Not Supported*
H3	When multitasking at their discretion, those who decide to switch more often will have lower accuracy than those who switch less often.	*Supported*
H4	Those forced to multitask during a task that they consider harder will perform worse than if they considered the task easier.	*Supported*
H5	When multitasking at their discretion, those who perceive the primary task as difficult will have more switches than those who consider the primary task not as difficult.	*Not Supported*
H6	Receiving external interruptions will lead to more anxiety, annoyance, and frustration.	*Not Supported*
H6a	Those forced to multitask will have more anxiety, annoyance, and frustration than those who do not multitask.	*Not Supported*
H6b	Those forced to multitask will have more anxiety, annoyance, and frustration than those who multitask at their discretion.	*Not Supported*

While some of the hypotheses were not supported, additional analyses were also done that revealed many different significant patterns. Within the discretionary condition only, an in depth analysis was performed and the results showed that the relationship between accuracy and multitasking is negative, where the more a subject multitasked the worse they performed. In terms of productivity, however, the results showed an inverted-U relationship where a little multitasking helped performance, but there was a point of diminishing returns where too much multitasking hurt performance.

Furthermore, these high multitaskers not only had a lower accuracy than all of the other discretionary multitaskers, but they also performed worse than all of the other subjects regardless of the condition. Middle multitaskers not only had a higher productivity than all of the other discretionary multitaskers, but they also performed better than all of the other subjects in any condition.

With regard to accuracy, those in the forced condition did perform significantly worse than most of those in the discretionary condition, with the exception as mentioned of the high multitaskers. Non-multitaskers tended to be divided in terms of those who had a propensity to multitask in general versus those who did not multitask in general. Surprisingly, subjects in the non-multitasking condition who had a propensity to multitask performed better, both in terms of accuracy and productivity, than subjects who did not have a propensity to multitask in the non-multitasking condition. In terms of accuracy, they also performed better than subjects in the forced condition and the high discretionary multitaskers. Non-multitasking subjects who did not have a propensity to multitask were not significantly different than those in the forced multitasking condition.

Additional discretionary analyses revealed different reasons why people multitasked. Reasons ranged from negative reasons, such as frustration, obstruction, and exhaustion, while positive reasons included boredom, exploration, and reorganization. Results showed that subjects in the negative condition had more switches and performed worse than subjects who switched for positive reasons.

CHAPTER 8. DISCUSSION

In the literature review chapter, under Norman's Stages of Execution, we discussed how multitasking individuals execute multiple action sequences corresponding to different tasks. In addition, we analyzed the drivers of multitasking and classified them into internal and external interruptions. The literature suggested that there is still plenty of opportunity to contribute to the field of Human-Computer Interaction (HCI) by undertaking more research on multitasking.

The aim of this dissertation was to study multitasking where previous research fell short. More specifically, this study sought to contribute to the HCI literature on multitasking by performing an empirical investigation comparing the performance of subjects who are forced to multitask due to external interruptions and subjects who decide on their own to multitask (i.e. forced and discretionary conditions). The study also examined the discretionary multitasking condition in more detail. The goal was to determine whether there were differences in terms of accuracy, productivity, and frustration between the different ways in which one multitasks or does not multitask. Additional goals included investigating whether subjects' perceived difficulty level of their primary task affected their accuracy in the different conditions or whether the level switching affected the performance for different users in the discretionary condition.

In order to answer these questions, a controlled experiment was conducted. For this experiment, a custom-environment in Microsoft Visual C++ was developed with three conditions: discretionary, forced, and the control group (non-multitaskers). The discretionary condition was tab-based, where each tab displayed a different task, and subjects were able to decide when and how often to switch tabs. In the forced multitasking condition while subjects were in the middle of their primary task a new pop-window with one of the secondary tasks would appear at certain times designated by the system (and determined by prior pilot tests). Subjects were forced to switch to and complete the interrupting task before being allowed to resume the primary task. In the non-multitasking condition, each task appeared sequentially. Once a subject's allotted time for the primary task was done, the secondary tasks would appear on the screen, one following the other.

A total of 636 subjects were recruited to participate in the experiment (212 in each condition). Their objective results, as recorded by the custom-developed environment and their subjective perceptions, captured by a post-test questionnaire were used to test the hypotheses put forth in Chapter 5.

While H1 predicted that subjects in the forced condition would have lower accuracy than subjects in the other conditions, the results do not support this prediction. Forced multitaskers did not have a worse accuracy than the other conditions. However when subjects felt the primary task was difficult, those in the forced condition had lower accuracy than those in both the discretionary and non-multitasking conditions. The harder a task was the more interruptions hurt their performance.

When a task was not considered difficult, subjects in the forced condition did not perform significantly worse than those in the two other conditions. However, when the task was perceived as difficult, the results were consistent with the first hypothesis. Receiving interruptions during a hard task can cause people to lose their train of thought and affect their performance. It appears that this is not the case for easier tasks where getting interrupted is less of a concern. During an easier task, a user does not have to spend as much time and energy recalling the details of the interrupted task. The prediction of H1 was only supported when the primary task was perceived as difficult.

In addition, subjects in the forced multitasking condition felt the second Sudoku puzzle was more difficult than subjects who did not multitask. This suggests that being forced to multitask during a difficult task can cause one to view the task as even more difficult.

In terms of productivity, while H2 predicted that non-multitaskers would perform better than discretionary and forced multitaskers, the results do not support this prediction. While it seemed as if the non-multitaskers should have been able to complete more of the tasks since they did not have to spend extra time recalling previous tasks, this was not the case. Compared to subjects in the forced or discretionary multitasking conditions, subjects who worked sequentially and did not multitask did not have higher productivity and therefore, H2 was not supported by the results.

One of the most surprising results in the non-multitasking condition was that performance was influenced by multitasking propensity. Non-multitaskers who had a higher propensity to multitask in general tended to perform better in terms of both accuracy and productivity than those with lower multitasking propensity. Perhaps since these multitasking individuals were used to performing many tasks, focusing on one task was simpler and therefore led to better performance.

In a purely discretionary condition, the results are consistent with the predictions of H3; subjects who multitasked more had lower accuracy. A more in-depth analysis revealed that high multitaskers performed worse than both low and medium multitaskers. However, in terms of productivity, middle multitaskers performed better than both low and high multitaskers. This implies that a little multitasking produces better productivity, but it is an illusion, since the greater the multitasking, the lower the accuracy. In fact, high multitaskers' accuracy was not only worse than all the other discretionary multitaskers, but all the other conditions as well.

With regard to productivity, though H2 was not supported and non-multitaskers did not perform the best, a detailed analysis with subjects in the discretionary condition does show some important patterns across conditions. Middle multitaskers not only performed better than both low and high multitaskers, but better than all the other conditions as well.

For subjects in the forced condition, it was expected that they would have lower accuracy if they considered the task to be harder rather than easier. The results supported the prediction of H4. In order to prove that when a task was considered harder subjects in the forced multitasking condition performed worse because of the interruptions they received rather than because the task was considered harder, the same analysis was done in the discretionary and non-multitasking conditions. Only in the forced multitasking condition subjects' accuracy was lower for harder tasks and higher for easier tasks. These results extend those obtained for H1. H1 compared accuracy across conditions and determined that subjects in the forced condition had lower accuracy than subjects in the other conditions when the primary task was considered difficult. H4, however, exclusively focused on the forced condition to determine whether receiving interruptions caused one's accuracy to be worse during tasks perceived as harder than during tasks perceived as easier. Taken together H1 and H4's results suggest that being forced to multitask during a difficult task can harm one's accuracy.

Challenging tasks can require extra concentration and receiving interruptions with secondary tasks can negatively affect accuracy. When performing under the pressure of interruptions, accuracy can suffer. Not only did accuracy suffer in the primary task, but for the secondary tasks as well. Only in the forced condition subjects who found the primary task difficult performed worse in the secondary tasks than subjects who found the primary task easier. This was not the case for subjects in the discretionary and non-multitasking conditions.

Participants in the discretionary condition were expected to switch more if they found the primary task difficult, compared to those that found the task easy (H5). Surprisingly, the results were significant but in the opposite direction. In the discretionary condition, when a subject found the Sudoku task difficult s/he tended to stay on task rather than switch away from it. This implies that the level of task difficulty did affect how often a subject switched. When a task is hard people can be engrossed in it and not want to be distracted. Sudoku is the type of puzzle where the answers are all connected with each other. The effects of losing concentration might be detrimental to performance. By remaining on the task, subjects can avoid losing their train of thought. The results of the pilot experiment showed that H5 was supported. It is possible that the different environments caused people to behave differently.

Furthermore, we previously saw that subjects who were frustrated or stuck on Sudoku had more switches. While this seems contradictory it is different. Switching out of frustration or clearing one's head can be different than finding a task difficult and therefore remaining on it. While subjects tended to stay on the task if they found it hard, qualitative analysis showed that when they were stuck, wanted to the clear their head, or were frustrated they had more switches.

It was expected that those in the forced multitasking condition would experience more anxiety, frustration, and annoyance than subjects in the discretionary or non-multitasking conditions (H6). The items to measure anxiety, frustration, and annoyance all loaded into one factor, since people's perceptions of these three aspects can be very similar. A new composite variable called frustration was created. The results of comparing this variable in the three conditions were not significantly different. Thus, H6 was not supported. It can be argued that perhaps it was not the environment that caused subjects to have more frustration or less frustration but their accuracy on the tasks themselves. A separate correlation analysis between the level of frustration with accuracy in each condition showed that the frustration of those in the discretionary or non-multitasking conditions did correlate with their accuracy. The lower their score, the more frustrated they were. This correlation was not significant for subjects in the forced condition. In general, the level of frustration was also influenced by the prior level of skills in the primary task and the level of computer skills; the lower their skills, the higher their frustration. While the condition did not appear to impact frustration level, other factors such as preparation for the task and the computer environment seem to play a bigger role.

8.1 Limitations

There are a number of limitations with this experiment. First, the subjects were college students, and the experiment was conducted in an academic environment. The results may not be applicable to different types of individuals or to other environments. An older population (or even a younger one) may differ in the amount of multitasking, their perception of multitasking, and propensity to multitask.

Furthermore, half of the subjects in the sample were paid $10 for their participation, while the other half received course credit. We saw how subjects in these two different incentive groups acted differently. It is possible that subjects in the paid condition tried to be more productive since they were paid and therefore had more switches, or their age or other demographic factors were variables explaining why they acted in a different way. Because of these differences, all the analyses include a control to indicate the type of incentive.

Although subjects received $10 or credit, they may not have had an incentive for trying their best and ensuring their optimal performance. However, the results of the data and informal interviews confirm that most of the participants took the tasks seriously. They found the multitasking environment engaging, challenging, and fun. For example, in the post-test questionnaire comment box, some of students' comments are: "The game was actually pretty fun and challenging", "This as a fun exercise", and "Tough but fun".

Another limitation is that the strict time controls built into the experimental environment may have affected multitasking patterns. In the open-ended question included in the post-test questionnaire for the discretionary condition, some subjects wrote that they did not multitask because of the timing. And in the comment box on the post-test questionnaire for all subjects regardless of the condition, many subjects expressed frustration with the time limits. In other situations, where time is not as strictly controlled as in this experiment, different results may be produced. In this case, however, time controls were essential to rule out the effects of time on performance.

Subjects have the ability to randomly put in answers which could affect both the productivity and accuracy scores. There is a warning box that pops-up when the completion time is approaching letting the subjects know how much time is left. Subjects can in theory enter any incorrect responses to complete the task at that time. To avoid this there was a message on each screen for the secondary tasks (since they intentionally do not have enough time to finish) indicating not to worry if they cannot finish on time (see Appendix G). In addition, in their instruction handout there is a message "Scores for the tasks are based on the number of responses completed, so don't worry if you do not have time to finish the tasks." Nevertheless, there is still the possibility that some subjects chose to enter random answers.

Another limitation of the study is that the accuracy measure for the primary task is not independent. Sudoku is the type of puzzle where one mistake can lead to many others. While the secondary tasks had questions or boxes to be filled in that were independent, one mistake in

Sudoku can produce more errors in the row, column, and 3 X 3 box, which can lead to errors throughout the entire Sudoku puzzle.

Another factor that could have contributed to performance was that subjects in the non-multitasking condition saw the secondary tasks after the 18 minutes of playing Sudoku. This could affect their performance since they may be more fatigued than forced multitaskers who experience the secondary tasks earlier or discretionary subjects who have the option of performing these secondary tasks earlier.

The type of tasks built into the experimental environment is also a source of limitations. In non-experimental laboratory settings, multiple tasks competing for the subject's attention may not be as straightforward as performing tasks all with correct or incorrect responses. Real-life tasks include decision-making, planning, generating new ideas and may not correspond to problem-solving (puzzle like) exercises. Nevertheless, having these tasks provided the means to objectively calculate performance scores for all subjects.

The design of this experiment did not take into account interruptions that occur outside the computer, such as the ringing of a phone or a person approaching to chat. Other common computer interruptions such as notifications of incoming emails were also excluded. Checking email occurs very frequently (Renaud et al. 2006), and the effects of that type of interruption may be different than being interrupted by (or switching to) a problem solving task where correct answers need to be produced. Real-life interruptions can also occur with more frequency and many times are related to each other. For example, having a back-and-forth instant message conversation while working on a project would be very different than the types of interruptions included in this experiment.

Although the limitations of this experiment (college students with timed tasks) suggest caution when generalizing the results to other populations and settings, the results provide the basis for analyzing the performance effects of multitasking behavior. By intentionally using tasks that can be "scored" (with the use of productivity and accuracy metrics) and time limits the results are comparable across subjects working under different conditions. The use of a laboratory experiment allowed for the controlling of the variables of interest (tasks and times) in order to test the hypotheses in the absence of extraneous influences. In addition, having the experiment solely on a computer in a controlled environment provided an effective way of automatically recording the subjects' activities in a log file. This allowed an accurate analysis to be performed which would not have been as successful if there were also other types of external, non-computer-based, types of interruptions.

8.2 Implications

8.2.1 Implications for Research

This study sought to contribute to the multitasking literature by addressing an important gap. On the one hand, there were very few studies discussing multitasking in the HCI literature from the perspective of the user. The majority of those studies examined notifications systems and interruptions and were mostly focused on improving systems and features. Consistent with this pattern, even fewer studies investigated discretionary computer-based multitasking. To the best of our knowledge, there has not been any comparing the performance between discretionary multitasking and multitasking due to interruptions. The objective of this research was to systematically compare the performance effects of different multitasking strategies.

This study provides a novel approach in that it compares the performance of those voluntarily choosing to multitask, those receiving interruptions and being forced to multitask, and those not multitasking. Previous research has found that receiving interruptions will cause a lower performance than when performing tasks sequentially (Bailey and Konstan 2006). The results of adding this third condition (discretionary multitasking) and comparing performance across all three was not significant. However, being forced to multitask when receiving interruptions (i.e. the forced multitasking condition) was the worst method when examining only those subjects who found the primary task more challenging.

Consistent with prior findings reported in the literature, this study also reports different effects depending upon the degree of difficulty of the task. Speier et al. (2003) found that interruptions hurt performance for complicated tasks and helped for simple tasks. This study considered difficult tasks as those tasks which subjects perceived as difficult and compared performance against the two other conditions. While there were no significant differences when a task was considered easy, when the task was considered harder, the subjects in the forced condition had the worst accuracy. In addition, only in the forced multitasking condition, accuracy in the primary and secondary tasks was significantly lower for those who found the primary task harder rather than easier. Therefore, researchers can continue to study how to minimize the negative effects of forced interruptions which can be helpful to the multitasking user.

Very little research to date has discussed performance in terms of users who multitask due to self-initiated switches, i.e. an internal reason causing them to multitask. Research is also sparse regarding the motivations for voluntarily switching tasks in the absence of interruptions. In this study, while subjects who had a medium amount of self-interruptions had the best productivity than any subject in any condition, subjects who had high levels of discretionary multitasking had the worst accuracy than subjects in any condition. Due to the negative effects of excessive discretionary multitasking on accuracy, it is important to understand what causes people to multitask in the first place. While there are not too many studies on discretionary multitasking, a few have discussed explanations for why people choose to multitask (Jin and Dabbish 2009; Payne et al. 2007). This research study has found six different reasons why people multitasked in this experiment and has divided them according to negative and positive reasons (three in each category). Those who multitasked for negative reasons tended to have the most switches and the worst accuracy. Consequently, HCI researchers can help by studying how to minimize the negative effects of interfaces in order to improve multitasking performance.

8.2.2 Implications for Interface Designs

Modern users typically perform multiple computer-based tasks at the same time. In modern personal computer environments, technological advances in computing power and notification capabilities are causing people to engage in multitasking. This brings up the challenge of creating better interfaces that make it easier for users to switch back and forth between projects (Czerwinski et al. 2004).

The goal of designing this interactive interface and testing these different multitasking conditions was to help HCI researchers gain more insight into how people multitask and how it affects their performance. In addition, the results of this experiment can also provide user interface guidelines for multitasking environments. HCI research is concerned with designing interfaces that can be more effective and efficient. The results of this research can help when determining how to design interfaces.

For instance, according to the results, when one multitasks very often performance is lower. Interfaces can be designed in such a way that multitasking is more efficient. For example, the activity-based environments, such as described in Bardram et al. (2006), mentioned in the literature review chapter, would be a great way to reduce the number of switches when people are voluntarily multitasking. In this case, one switch would bring all the necessary applications to the foreground without having to do unnecessary switching. Sometimes when bringing up a new application for the purpose of one task, a user could get sidetracked by other windows or applications that are open. If computer systems were designed in such a way that one click would bring about all applications associated with one task this would be much more efficient and cause less switching away from a task. Voida and Mynatt (2009) describe an activity-based system that also has tagging capabilities. Tasks tend to be collaborative with other tasks and some may share certain files. A tagging capability may be a good feature that this type of system could have.

The idea of clustering related tasks has been proposed in specific contexts. For example, Scupelli et al.'s (2005) instant messenger application grouped contacts together in terms of tasks. It is more understandable and saves time when arranged in this way. When designing new interfaces, it would help to think of the goal of the overall task that is intended to be accomplished and to design programs in terms of tasks rather than applications. Thinking of designs in terms of the broader purpose of a task rather than an application can help in making it more efficient with fewer switches.

Qualitative analysis for the discretionary multitasking condition revealed different reasons why a subject chose to multitask. One of the findings was that people multitask for positive reasons (boredom, exploration, and reorganization) or negative reasons (frustration, obstruction, exhaustion). While these responses were provided for this particular experiment it can be extended to why people multitask in general. Whether one is working on a project for school or work or even leisure, oftentimes s/he can become bored, frustrated, exhausted or need to clear his/her head. A person may also desire to first view or respond to smaller tasks, like checking their email or social networking site. Therefore better interfaces can be designed when knowing the reasons why people tend to multitask.

For example, according to the results provided in this study, when a user is switching for negative reasons, such as frustration or when one is stuck, s/he tends to have more switches and worse performance. Therefore, creating simpler interfaces that lead to less confusion or frustration can prevent a user from wanting to switch. The simpler the design is the better. Subjects had less switches when they switched for positive reasons, such as curiosity or being bored, rather than when they were frustrated or stuck and did not know how to proceed. In those cases the subject's accuracy was worse as well.

Forced multitaskers had lower accuracy than both discretionary multitaskers and non-multitaskers when a task was considered difficult, otherwise there was no difference in terms of accuracy. This implies that in a difficult task, or more accurately a task that the subject perceives as difficult, receiving an IM alert or an email notification can hurt performance. Perhaps systems can be designed with a simple checkbox setting that subjects can click on when they are in the middle of a more complicated task. This would tell the system not to allow any environmental, external interruptions.

8.2.3 Implications for Practice

The results of this study also have practical implication for information workers in organizations and computers users in general. In practice, for challenging tasks, multitasking may not be the best strategy. When receiving an interruption during a hard task, there may be negative consequences to a person's performance. In contrast in an easier task, getting interrupted did not seem to affect a user's performance. In terms of voluntarily multitasking, too much switching, regardless of the level of difficulty of the task, may hurt one's accuracy. The effects are worse than if the task had been performed sequentially or in any other method.

Keeping these insights in mind will allow one's work to be as effective as possible. Palladino (2007) discusses mindful multitasking and how people should be aware of their own multitasking limits, since too much multitasking will hurt their performance. The results of this experiment suggest that the more tasks one is completing simultaneously the worse their accuracy may be. Therefore, if a task needs to get done correctly multitasking is not the best solution. Multitaskers should avoid excessive multitasking or their accuracy may degrade to the point where it suffers too much. While multitaskers may get more tasks completed, accuracy will be lower. Completing more tasks at a faster rate is not a great accomplishment, if the accuracy for those tasks is lower.

8.3 Future Research Directions

8.3.1 Additional Tasks

Examining more secondary tasks may provide an extension to this experiment. More tasks would allow for more interruptions while the primary task is completed. For example, having a back and forth instant message conversation would require more task switches. Future work can vary the number of tasks, so that the number of tasks is the independent variable with performance as the dependent variable.

This approach would be similar to how the discretionary condition examined the number of switches with accuracy and productivity and found that middle multitaskers performed better in terms of productivity, it would be interesting to see if this pattern held true in a forced condition where tasks are varied. For instance, would the number of different tasks show a similar to pattern that was found with the number of switches? Would forced multitaskers who receive a medium amount of interrupting tasks have the best productivity, and forced multitaskers who receive the most interrupting tasks have the worst accuracy than subjects who receive a medium or smaller amount of interrupting tasks? Future work can vary the number of tasks in order to study in more detail the forced multitasking condition as was done in this experiment with the discretionary multitasking condition.

8.3.2 Self-Initiated Tasks

Another potential extension for future research is to allow subjects to initiate some of their own tasks, such as browsing the web or checking their email. Allowing the subject to have a little more flexibility with regard to how they would normally use the computer, may prove insightful. A subject's performance can be examined if some of the secondary tasks were initiated by the subject him/herself.

8.3.3 Different Primary Task

Sudoku experience played a large role in this study. Many times when examining the control variables on the difference ANOVAs conducted, Sudoku experience is significant.

Subjects who had more previous experience with Sudoku had a greater accuracy than subjects who had less experience. One way to minimize this effect would be switch the primary task to a different task, one where there is a more equal level of skill set, i.e. a task more subjects are familiar with or a completely new task which fewer subjects are familiar with. While the practice round tried to minimize some of the differences between those who never played and played Sudoku before, one cannot really be experienced in playing Sudoku until they have played it at least a few times.

8.3.4 Timing

A future experiment can have secondary tasks with longer time limits. This may prevent subjects from being afraid to multitask because of the short time duration.

8.3.5 Non-Laboratory Experiment

Another possibility is to study multitasking in a field setting (i.e. workplace, home, or school) by observing or interviewing subjects about their multitasking habits. Although assessing performance would be more difficult in a field setting, the results in terms of multitasking patterns can be compared to those obtained in this experiment.

8.3.6 Non-Computer Based Multitasking

Future research can analyze the effects of performance when one is not solely multitasking on the computer. Additional non-computer based tasks can be added. Receiving a phone call or chatting with a coworker, friend, or colleague can impact performance.

Furthermore, other multitasking platforms can be examined. One such example would be the iPhone. Technology is constantly evolving. Products are becoming faster, cheaper, and including more space for many more user applications. Multitasking is therefore becoming more and more prevalent and additional effort in studying multitasking can help determine what triggers people to multitask, how multitasking affects performance, and how interfaces can be designed to account for the modern multitasking user.

8.3.7 Study Design Mechanisms

Future research can study different design mechanisms in order to create multitasking interfaces that improve both accuracy and productivity. Systems can have two different attributes: *restrictiveness* or *guidance* (Silver 1990). While restrictiveness pertains to systems that can limit a user to only certain functionality, guidance refers to systems that guide users into choosing certain decisions.

In the interface used in this experiment, both the non-multitasking condition and forced multitasking condition were restrictive in that users were only allowed to perform tasks in the order designated by the system itself. While the discretionary condition appeared to have no limits, it did guide the subjects by initially starting them at Sudoku and showing a timer on that screen. This could have guided subjects to remain on Sudoku when normally they tend to multitask.

To improve the effects of multitasking on performance, systems can be designed in such a way that they guide users in effective ways to multitask. Rather than having systems that limit multitasking in a restrictive manner or enable people to multitask solely at their discretion, systems can offer guidance. This can be done in a few different scenarios: users can be directed to their primary task at the start of a new computer session or reminders can be set alerting users at certain intervals that their primary task is still open and has not yet been completed. Future research can try out different design mechanisms in order to test whether different guidance techniques can help improve the performance of multitaskers.

CHAPTER 9. CONCLUSION

This dissertation began by performing an in-depth review of multitasking literature from the perspective of HCI. Multitasking is very common in the home, workplace, and in the classroom. Multitasking occurs either due to external interruptions, such as receiving a notification which causes one to switch to another task, or to an internal interruption, where the user decides to multitask, such as when a user becomes bored or frustrated with the current task. Whether people multitask for external or internal reasons, different interfaces can be designed in order to help these two types of multitasking individuals.

In order to further understand the effects that computer-based multitasking has on performance, a laboratory experiment was created to test the effects multitasking has both when internally or externally multitasking.

This experiment compared the performance of three different conditions: discretionary multitasking (those allowed to choose whether and/or when to multitask), forced multitasking (those forced to multitask at allocated times designated by the system), and non-multitasking (those who did not multitask). The goal of this study was to answer three questions:

1. Are there differences in performance (speed and accuracy) in the primary and interrupting tasks within these three different conditions?
2. Are the effects of interruptions different for tasks that subjects find easier versus tasks that they consider harder?
3. How often do people switch when able to multitask at their discretion and how does that affect their performance?

The results have shown that for those subjects who considered the primary task difficult, those in the forced multitasking condition had the lowest accuracy scores. In addition, only in the forced condition did subjects who found the primary task difficult have a lower accuracy than subjects who considered the primary task easier. Therefore, receiving interruptions during a hard task had negative consequences.

When the level of difficulty was not accounted for, then high discretionary multitaskers had the worst accuracy while medium multitaskers had the best productivity. This implies that when people multitask they may feel that it is a better approach than working sequentially since they are accomplishing more, however in terms of the accuracy, it is not better, but may in fact be worse. In the discretionary condition the more switches one had correlated with a lower performance.

While it may seem that if a subject found a task difficult they would switch more often, in fact, the opposite was true. Subjects who found a task difficult tended to have less switches away from the primary task and actually tended to stay on the task. This suggests that the level of difficulty of a task also impacts the way people multitask.

This study also explored why people multitask and the impact the different reasons for multitasking has on performance. The results show that those who switched for negative reasons, such as becoming frustrated or stuck, had the most switches and the lowest performance.

The goal of this multitasking study was to educate designers of applications on how current users multitask and how their performance is affected by multitasking. This can enable the design of systems that are more efficient and effective.

Appendix A

14 Articles on multitasking from the following HCI journals:

ACM Transactions on Computer-Human Interaction (TOCHI)
International Journal of Human Computer Interaction (IJHCI)
International Journal of Human-Computer Studies (IJHCS)
Human Computer Interaction (HCI)
Interacting with Computers
Behaviour and Information Technology (BIT)

Author(s),Year	Research Question/Objective	Journal
Bailey and Iqbal (2008)	Examine how a user's mental workload changes during task execution.	TOCHI
Bellotti et al. (2005)	Examine the phenomenon of email overload.	HCI
Crook and Barrowcliff (2001)	What are the patterns of home computer usage among college students?	IJHCI
Johnson et al. (2003)	How can we reduce our limited understanding of collaborative and multiple tasks?	TOCHI
McCrickard et al. (2003a)	Evaluate Notification Systems.	IJHCS
McCrickard et al. (2003b)	How to evaluate notification systems.	TOCHI
McCrickard et al. (2003c)	How to optimize the user's experience when being interrupted.	IJHCS
McFarlane (2002)	How to handle interruptions.	HCI
McFarlane and Latorella (2002)	Why interruption is an important HCI problem, and will continue to grow in importance.	HCI
Oulasvirta and Saariluoma (2004)	Whether interruptions affect memory accuracy.	BIT
Oulasvirta and Saariluoma (2006)	How to overcome interruptions without adverse effects.	IJHCS
Renaud et al. (2006)	Explores the effects of email.	IJHCI
Trafton et al. (2003)	Examine the effects of having preparation time prior to an interruption.	IJHCS
Zhang et al. (2005)	What are the differing control strategies and performance of monochronic vs. polychronic individuals?	IJHCS

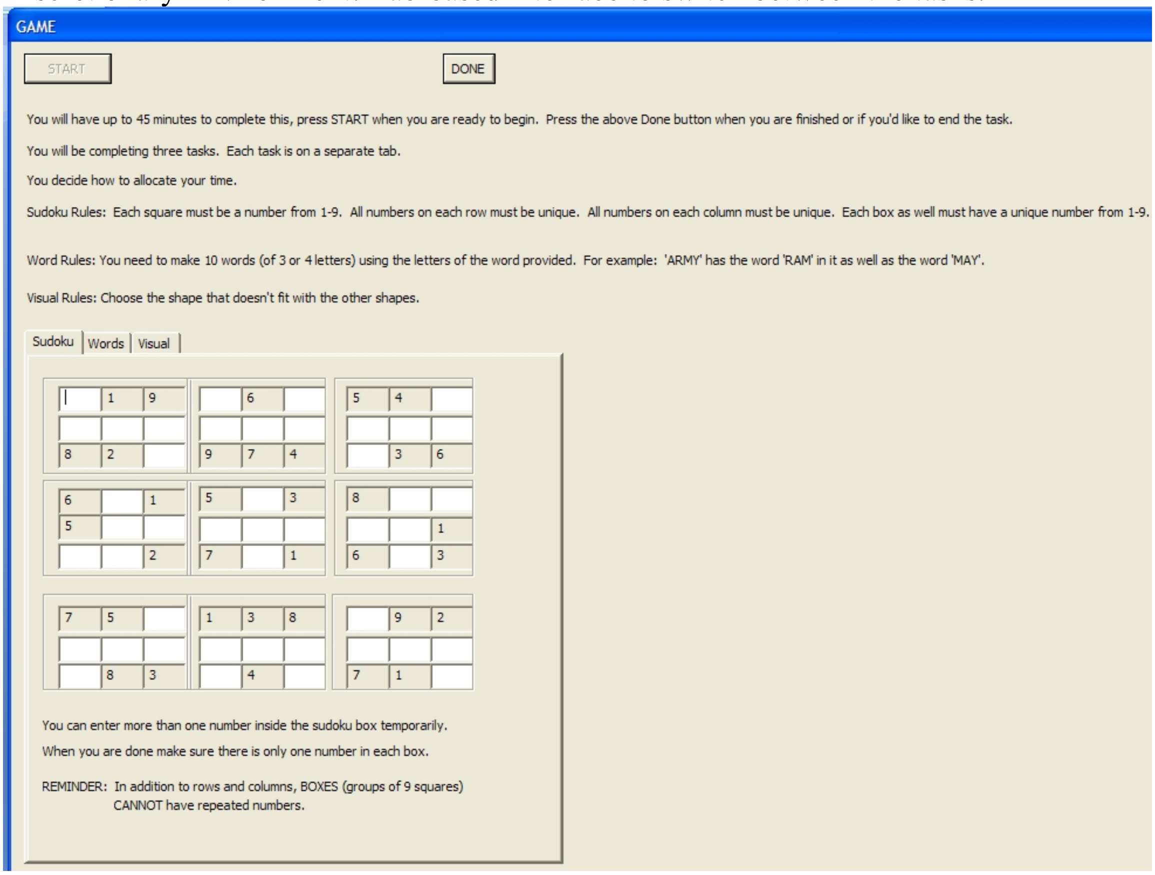
Appendix B

22 articles on multitasking from CHI proceedings.

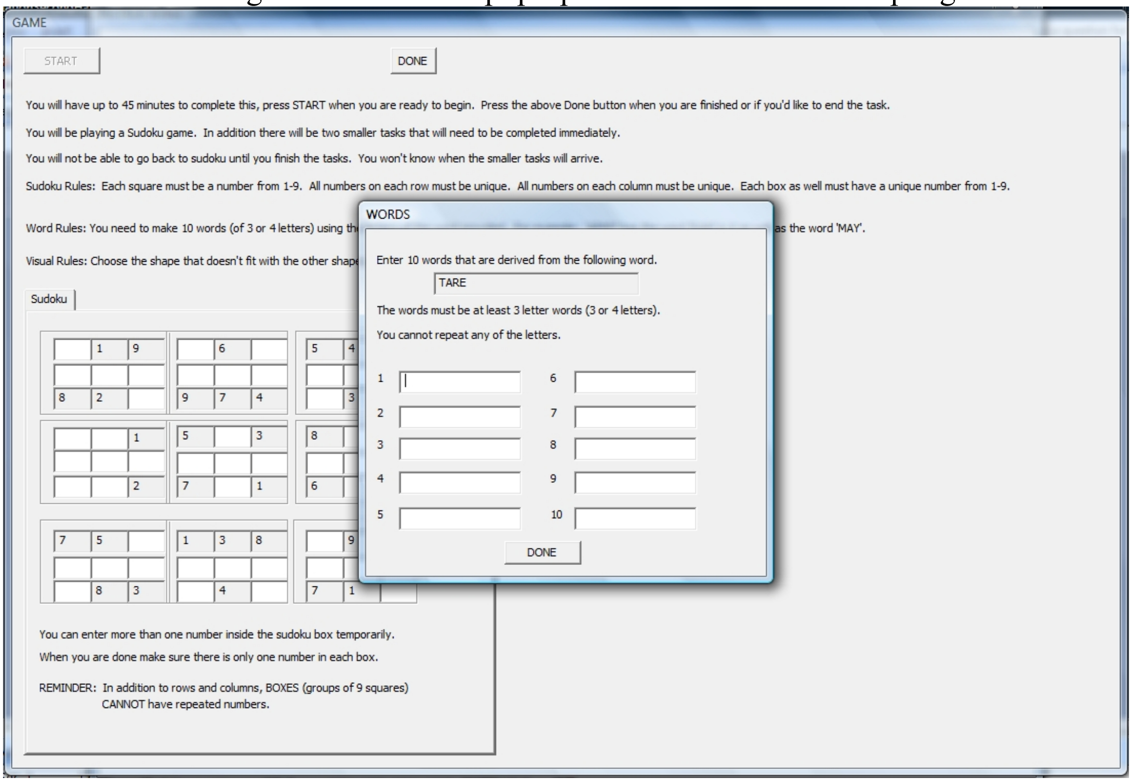
Author(s),Year	Research Question/Objective
Bardram et al. (2006)	Using activity-based computing (ABC).
Benbunan-Fich et al. (2009b)	Propose new metrics to investigate computer-based multitasking behavior.
Card and Henderson (1987)	Design an interface to help users task switch.
Czerwinski et al. (2004)	Examined people interleaving multiple tasks when interrupted.
Czerwinski et al. (2006)	Evaluate usability issues for large displays.
Gonzalez and Mark (2004)	To understand how information workers manage multiple activities.
Iqbal and Bailey (2006)	Examines how well characteristics of task structure predict Cost Of Interruption (COI), as measured by resumption lag.
Iqbal and Bailey (2007)	Detecting breakpoints in tasks.
Iqbal and Bailey (2008)	Present a Notification Management System
Iqbal and Horvitz (2007b)	Examine the effects of interruption on task switching.
Isaacs et al. (2002)	How people use IM
Jin and Dabbish (2009)	Examine internal interruptions.
Karam and Schraefel (2005)	Using gestures to support secondary tasks.
Kleinman (2009)	How are face-to-face meetings impacted by technological multitasking?
Mark et al. (2005)	Examine fragmented work.
Mark et al. (2008)	Does the context of interruptions make a difference?
Matthews et al. (2006)	Techniques to improve multitasking. They compare semantic content extraction (displays only the most relevant content) and change detection (signals when a change has occurred in a window).
Salvucci et al. (2009)	Unified theory of multitasking.
Scupelli et al. (2005)	Using IM to juggle many projects and teams.
Su and Mark (2008)	Focusing on communication chains to better understand multitasking.
Voida and Mynatt (2009)	Study activity-based computing as opposed to the application- and document-centric of most desktop computing interfaces.
Weisz et al. (2007)	Examine chatting while watching a video.

Appendix C

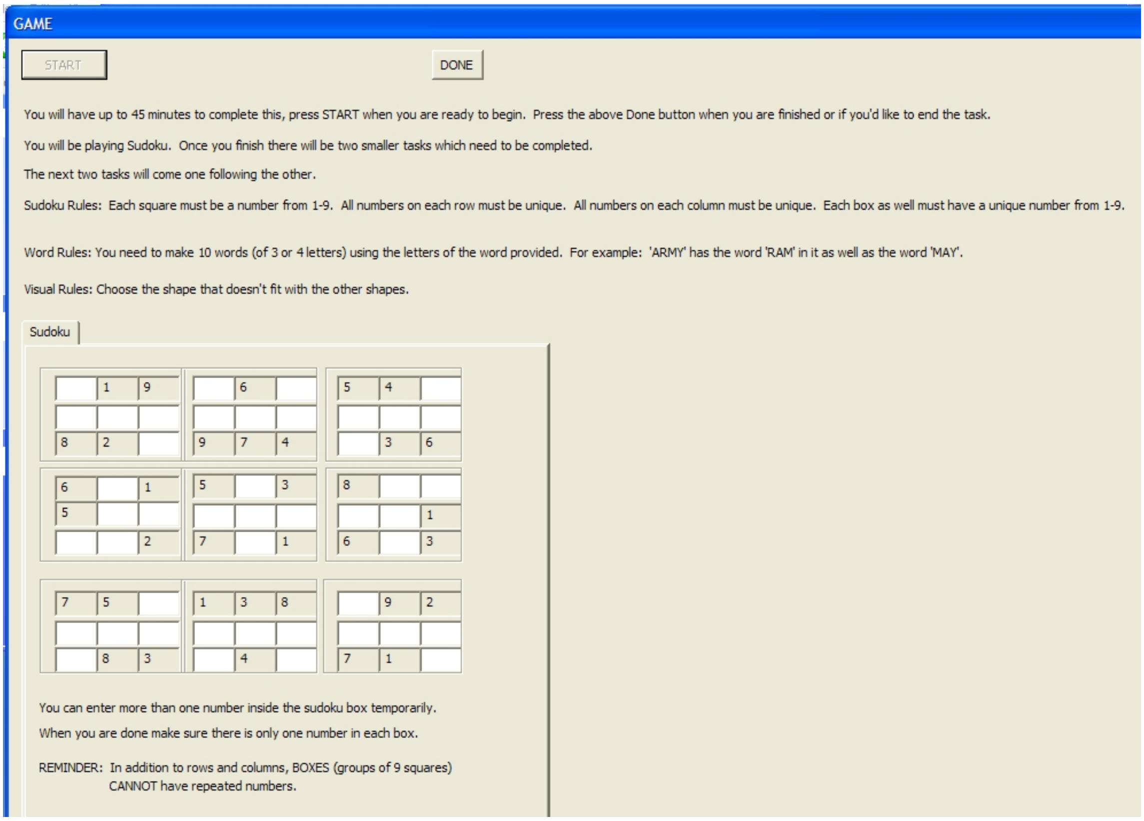
Discretionary Environment: Tab-based interface to switch between the tasks.



Forced Multitasking Environment: A pop-up window with the interrupting task.



Non-Multitasking Environment: Once the first task is completed, a pop-up window will appear with the next task.



Appendix D

Approved by
Baruch IRB
for use through
1/6/10

Appendix A – Informed Consent Form

Project Title: Performance Effects of Computer-Based Multitasking Behavior

Principal Investigators:
Raquel Benbunan-Fich (Baruch College) and Rachel Adler (Department of Computer Science, CUNY Graduate Center)

Informed Consent Form

The primary purpose of this study is to examine the performance of users on multiple unrelated computer-based tasks. Participants will be asked to complete three tasks. By participating in this study, you will help advance research, get paid for your time, as well as have fun playing games. We anticipate many sessions that will take place and require about an hour per session. You can only participate in one session.

You do not need to bring or have anything to participate in the study. All instructions will be explained before the actual experiment. Participation is completely voluntary. You may discontinue participation at any time. If you decide to participate, you will receive $10 for your time.

The risks from participating in this study are no more than the risk when playing any computer-based game. The benefit of your participation is enjoyment and financial gain. In addition, you will help enhance research in human-computer interaction.

All records of this study will be kept confidential. No one other than the PI and Co-PI (Raquel Benbunan-Fich and Rachel Adler) will have access to the data, which will be archived in a secure location after use. If desired, you may see the data collected from only your participation. Any resulting publications from this study will not identify individual participants but will refer to aggregate results.

If you have any questions regarding this research, you can call Raquel Benbunan-Fich at 646-312-3375 or rbfich@baruch.cuny.edu or Rachel Adler at 347-528-4067 or rachelfadler@gmail.com. If you have any questions concerning your rights as a participant in this study, you can contact Keisha Peterson, IRB Administrator, Baruch College/City University of New York at (646)-312-3785.

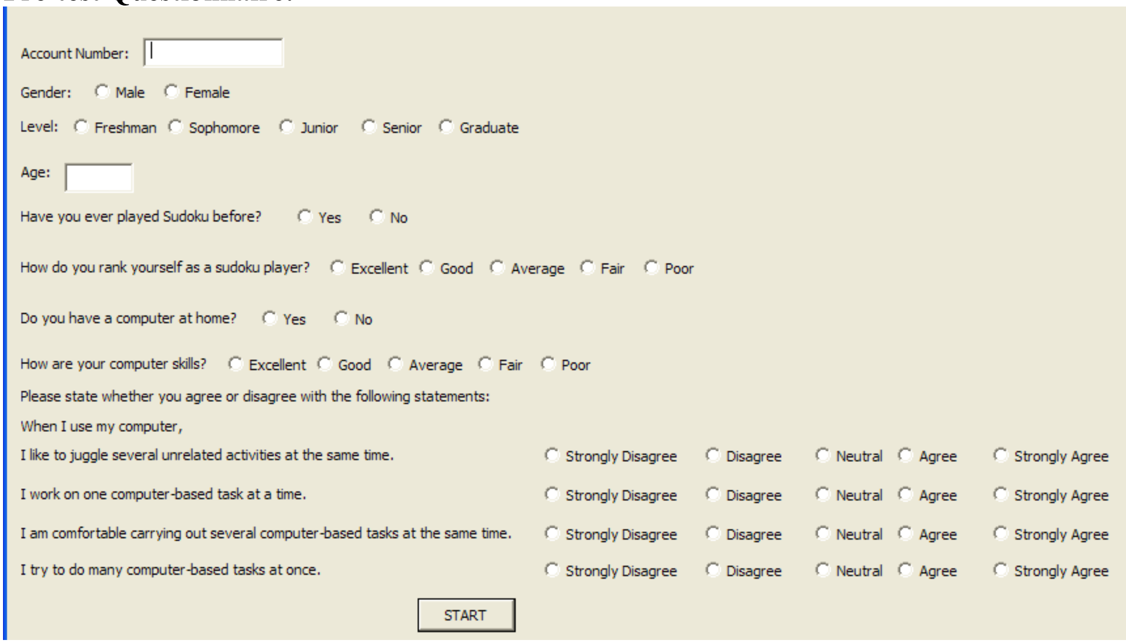
By signing below, you understand and accept the terms of this research study as stated above and that your participation is completely voluntary.

____________________________ ________________________
Student Signature Investigator's Signature

____________________________ ________________________
Full name (please print) Date

Appendix E

Pre-test Questionnaire:



Account Number:

Gender: ○ Male ○ Female

Level: ○ Freshman ○ Sophomore ○ Junior ○ Senior ○ Graduate

Age:

Have you ever played Sudoku before? ○ Yes ○ No

How do you rank yourself as a sudoku player? ○ Excellent ○ Good ○ Average ○ Fair ○ Poor

Do you have a computer at home? ○ Yes ○ No

How are your computer skills? ○ Excellent ○ Good ○ Average ○ Fair ○ Poor

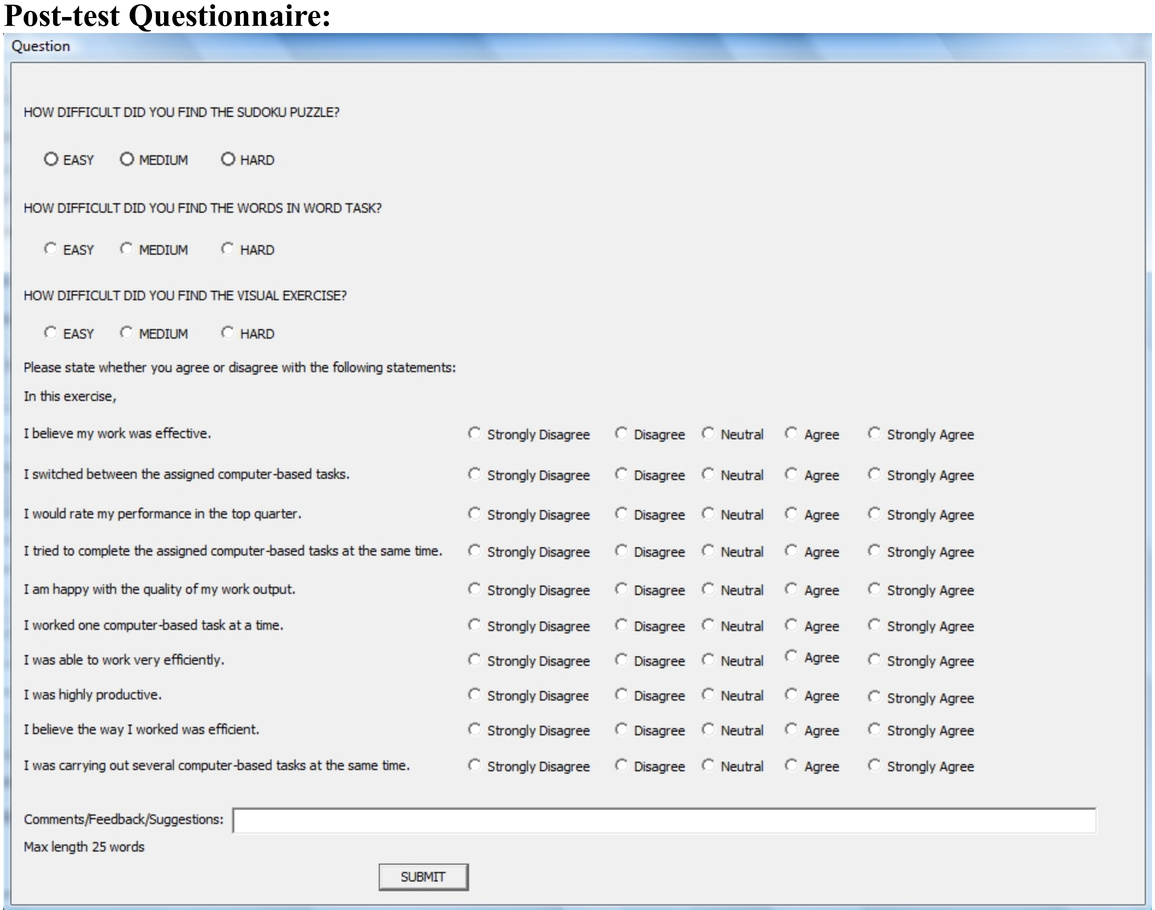
Please state whether you agree or disagree with the following statements:

When I use my computer,

I like to juggle several unrelated activities at the same time.	○ Strongly Disagree	○ Disagree	○ Neutral	○ Agree	○ Strongly Agree
I work on one computer-based task at a time.	○ Strongly Disagree	○ Disagree	○ Neutral	○ Agree	○ Strongly Agree
I am comfortable carrying out several computer-based tasks at the same time.	○ Strongly Disagree	○ Disagree	○ Neutral	○ Agree	○ Strongly Agree
I try to do many computer-based tasks at once.	○ Strongly Disagree	○ Disagree	○ Neutral	○ Agree	○ Strongly Agree

START

Post-test Questionnaire:



Question

HOW DIFFICULT DID YOU FIND THE SUDOKU PUZZLE?

○ EASY ○ MEDIUM ○ HARD

HOW DIFFICULT DID YOU FIND THE WORDS IN WORD TASK?

○ EASY ○ MEDIUM ○ HARD

HOW DIFFICULT DID YOU FIND THE VISUAL EXERCISE?

○ EASY ○ MEDIUM ○ HARD

Please state whether you agree or disagree with the following statements:

In this exercise,

I believe my work was effective.	○ Strongly Disagree	○ Disagree	○ Neutral	○ Agree	○ Strongly Agree
I switched between the assigned computer-based tasks.	○ Strongly Disagree	○ Disagree	○ Neutral	○ Agree	○ Strongly Agree
I would rate my performance in the top quarter.	○ Strongly Disagree	○ Disagree	○ Neutral	○ Agree	○ Strongly Agree
I tried to complete the assigned computer-based tasks at the same time.	○ Strongly Disagree	○ Disagree	○ Neutral	○ Agree	○ Strongly Agree
I am happy with the quality of my work output.	○ Strongly Disagree	○ Disagree	○ Neutral	○ Agree	○ Strongly Agree
I worked one computer-based task at a time.	○ Strongly Disagree	○ Disagree	○ Neutral	○ Agree	○ Strongly Agree
I was able to work very efficiently.	○ Strongly Disagree	○ Disagree	○ Neutral	○ Agree	○ Strongly Agree
I was highly productive.	○ Strongly Disagree	○ Disagree	○ Neutral	○ Agree	○ Strongly Agree
I believe the way I worked was efficient.	○ Strongly Disagree	○ Disagree	○ Neutral	○ Agree	○ Strongly Agree
I was carrying out several computer-based tasks at the same time.	○ Strongly Disagree	○ Disagree	○ Neutral	○ Agree	○ Strongly Agree

Comments/Feedback/Suggestions:

Max length 25 words

SUBMIT

Appendix F

Game Instructions

You will be asked to complete three tasks. Instructions will be provided letting you know the order of the tasks. You will have up to 45 minutes to complete all the tasks. Do not hit the done button, until you are sure you are done with the task.

TASK 1: Sudoku Rules: The goal of a Sudoku puzzle is to fill in all the boxes in a 9 X 9 grid, so that each column, each row, and each of the nine 3 X 3 boxes have the numbers 1 – 9 without any of those numbers being repeated.

- Each square must be a number from 1-9.
- All numbers on each row must be unique.
- All numbers on each column must be unique.
- Each 3 X 3 box as well must have a unique number from 1-9.

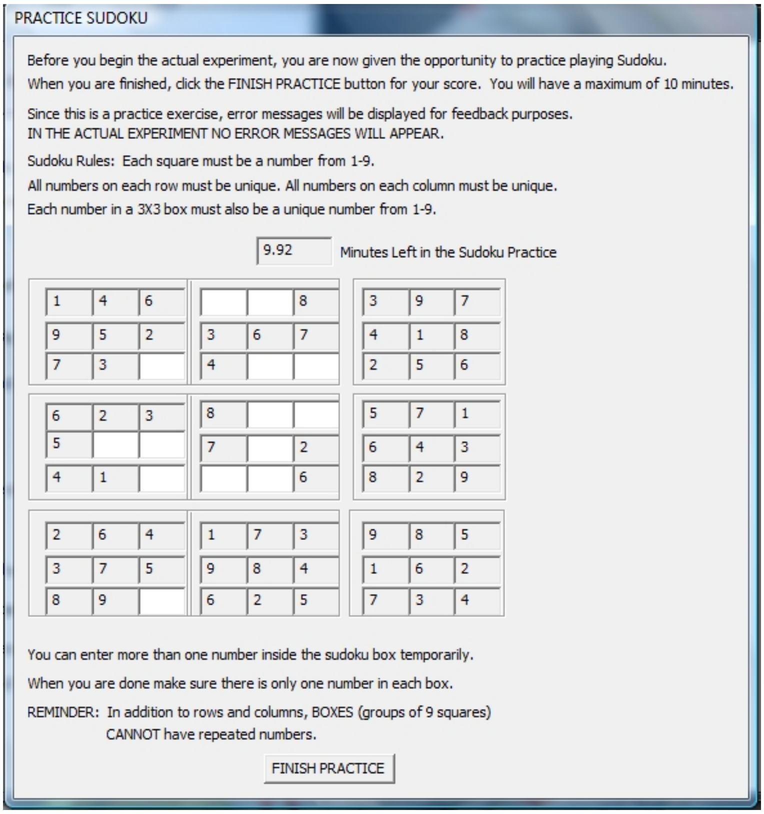
In the Sudoku, you can temporarily enter more than one number in a square to help you narrow it down. Remember to remove the extra numbers before submitting your answer.

TASK 2: Word Rules: You need to make 10 words (of 3 or 4 letters) using the letters of the word provided. For example: If the word was 'ARMY', it has the word 'RAM' in it as well as the word 'MAY'.

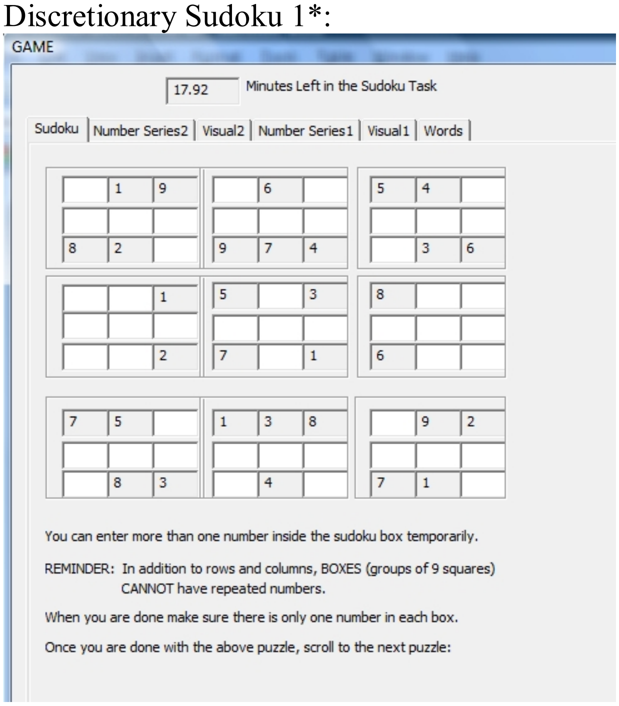
TASK 3: Visual Rules: Choose the shape that doesn't fit the same pattern as the other shapes.

Appendix G

Practice Sudoku Round:

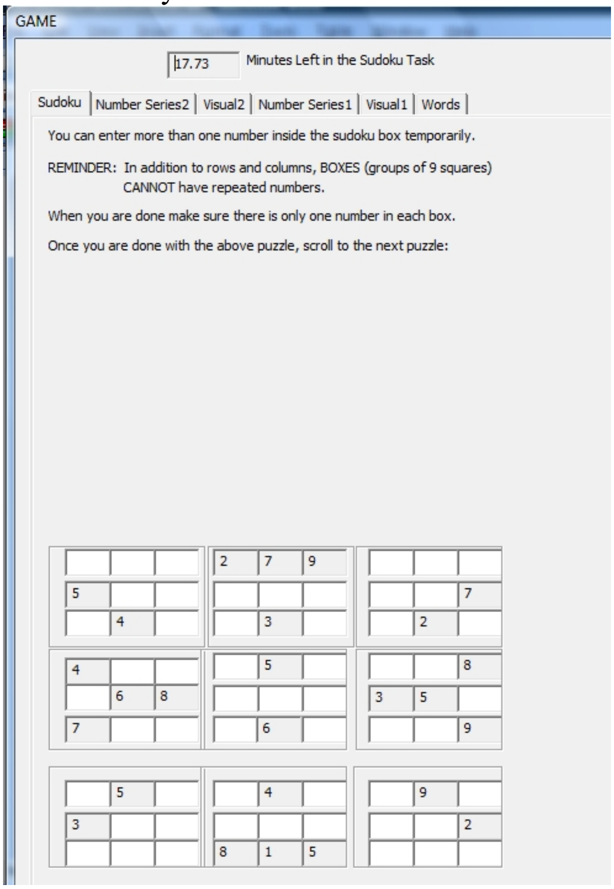


Discretionary Sudoku 1*:



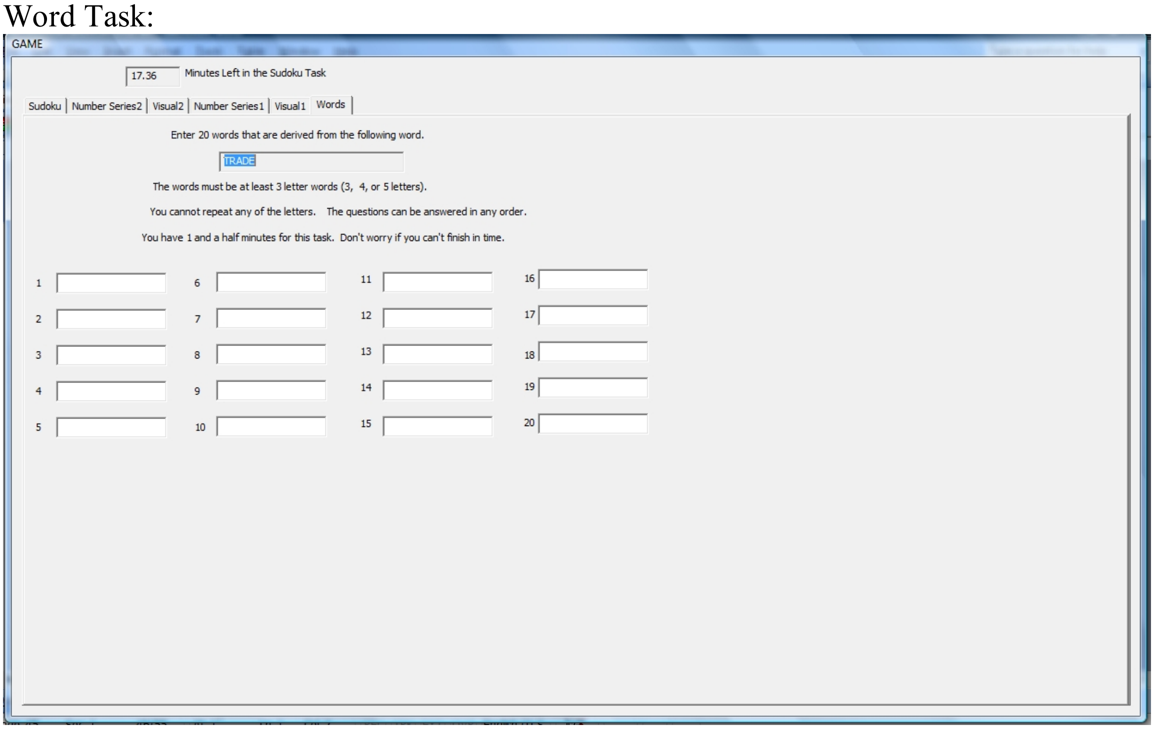
*This figure is cropped and enlarged.

Discretionary Sudoku 2*:

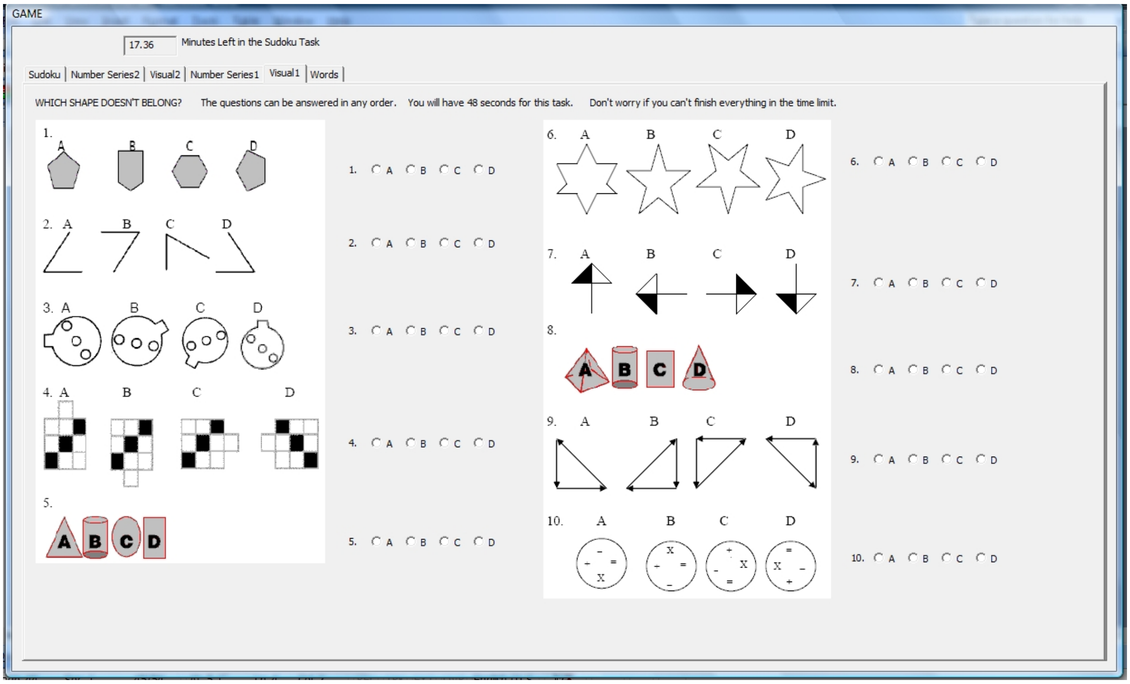


*This figure is cropped and enlarged.

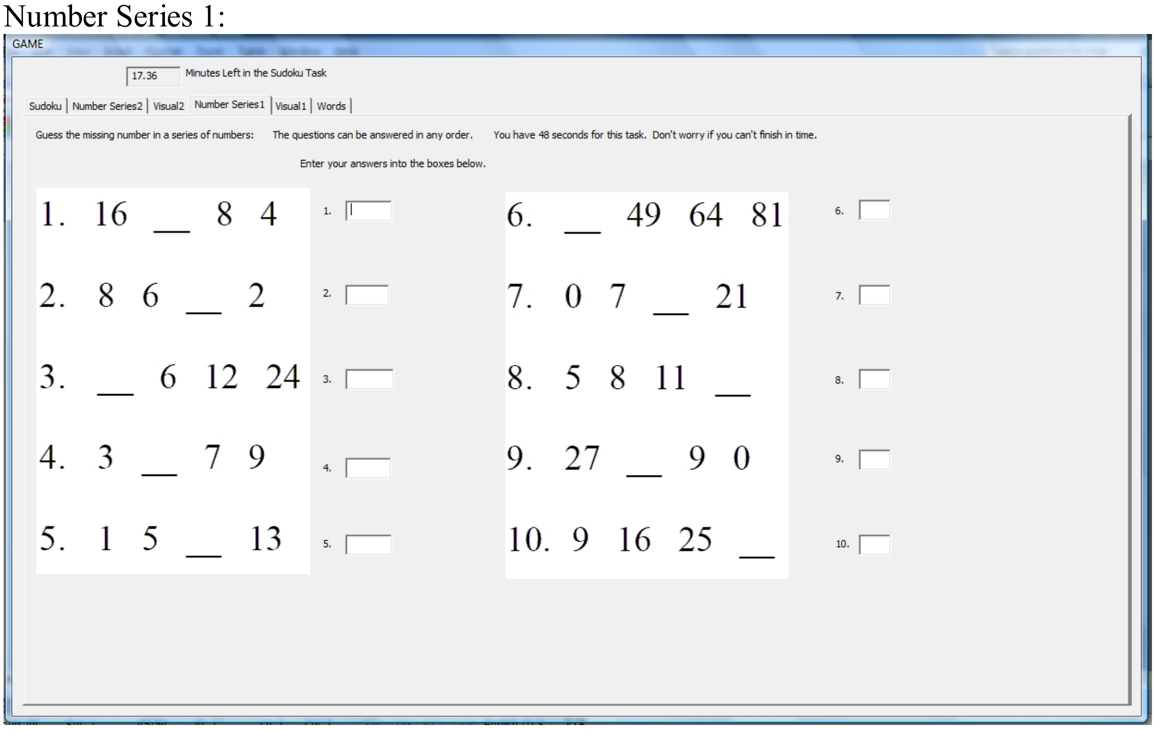
Word Task:



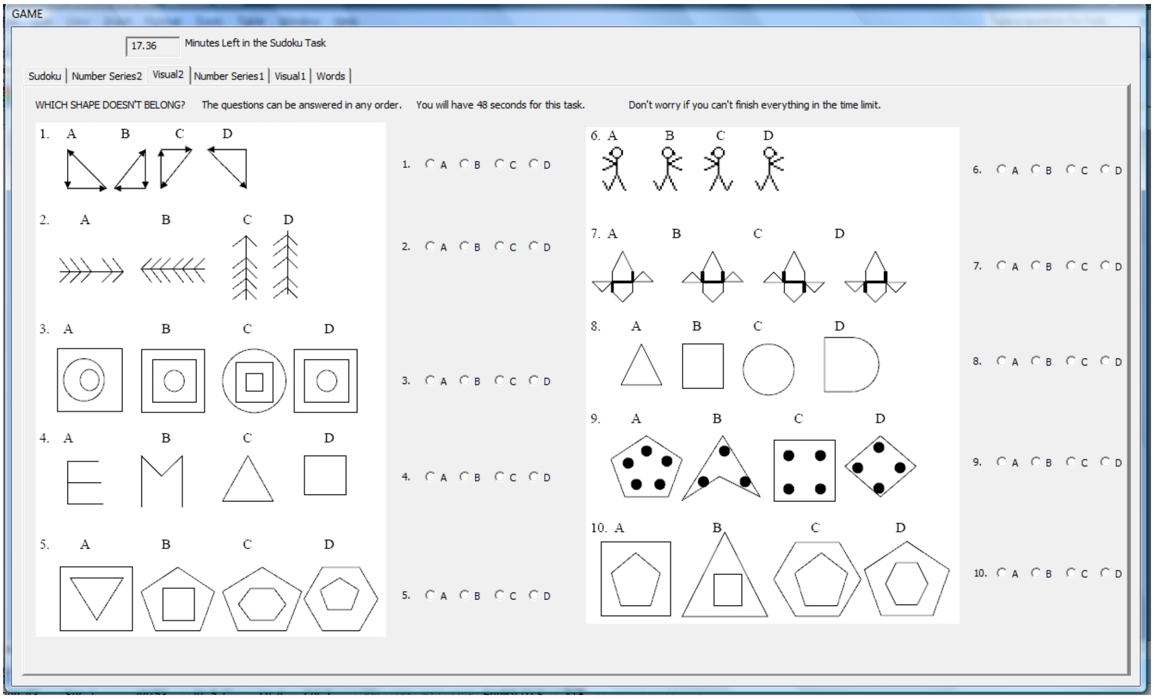
Visual 1:



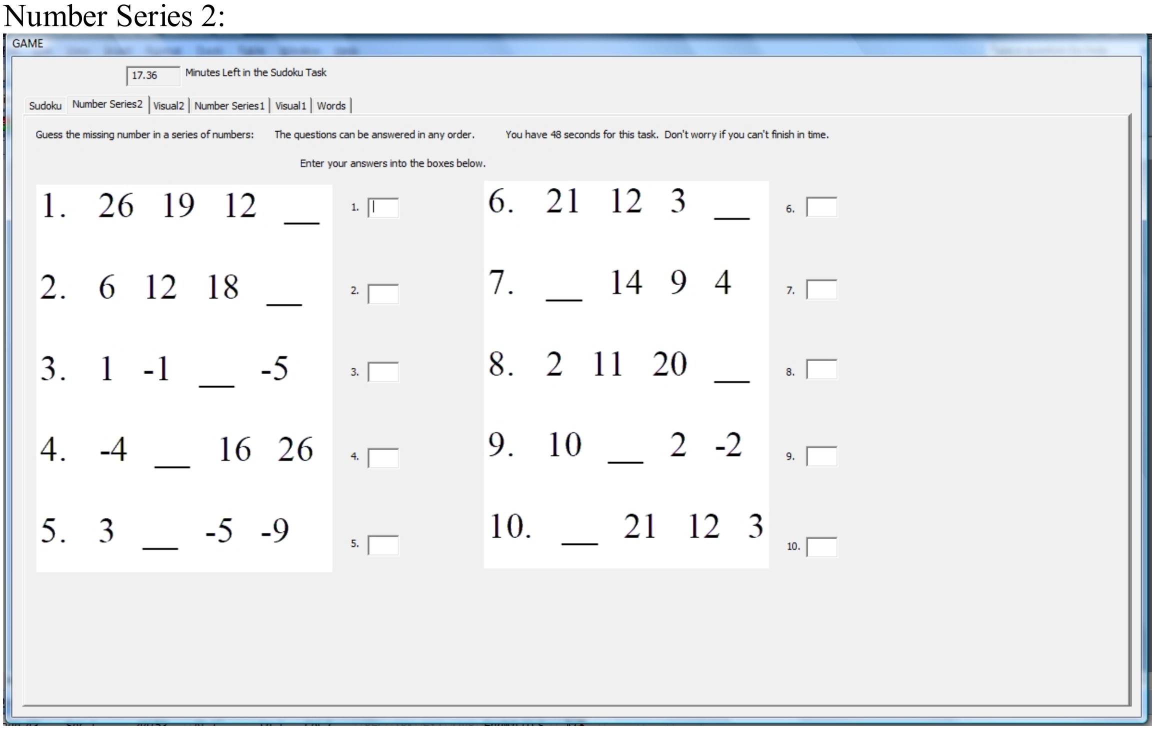
Number Series 1:



Visual 2:



Number Series 2:



Appendix H

Game Instructions – Non-Multitasking Condition

You will be completing a Sudoku puzzle, along with five smaller tasks. If you finish Sudoku early there is a **SECOND SUDOKU** puzzle below it. After you are done with Sudoku, you will complete the five smaller tasks one following the other.

Before you begin the game, a practice session of Sudoku will appear. You will have at most 10 minutes to practice. Since this is a practice exercise, error messages will be displayed for feedback purposes as you enter numbers into the squares. PLEASE NOTE: IN THE ACTUAL EXPERIMENT NO ERROR MESSAGES WILL APPEAR.

Scores for the tasks are based on the number of responses completed, so don't worry if you do not have time to finish the tasks.

The system will keep track of time. You will have 18 minutes for Sudoku, 1 and a half minutes for the word task, 48 seconds for each of the two Visual Tasks, and 48 seconds for each of the two Number Series Tasks.

The smaller tasks might be presented in a different order from the ones listed here:

Sudoku Rules: The goal of a Sudoku puzzle is to fill in all the boxes in a 9 X 9 grid, so that each column, each row, and each of the nine 3 X 3 boxes have the numbers 1 – 9 without any of those numbers being repeated.

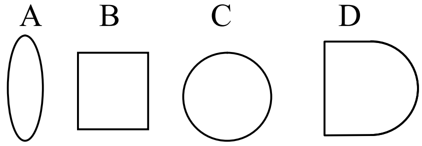
- Each square must be a number from 1-9.
- All numbers on each row must be unique.
- All numbers on each column must be unique.
- Each 3 X 3 box as well must have a unique number from 1-9.

In the Sudoku, you can temporarily enter more than one number in a square to help you narrow it down. Remember to remove the extra numbers before submitting your answer.

Word Rules: You need to make 20 words (of 3, 4, or 5 letters) using the letters of the word provided. For example: If the word was 'ARMY', it has the word 'RAM' in it as well as the word 'MAY'.

Visual Rules: "Odd One Out" Choose the shape that doesn't fit the same pattern as the other shapes. Example:

A B C D



B would be the correct answer, since B is the only shape without any round edges.

Number Series Rules: Guess the missing number in a series of numbers.
Example: 2 4 8 __ The missing number is 16, since each number in the series is doubled.

Game Instructions – Forced Multitasking Condition

You will be completing a Sudoku puzzle, along with five smaller tasks. If you finish Sudoku early there is a **SECOND SUDOKU** puzzle below it. **The five smaller tasks will appear while you are playing Sudoku, and will need to be completed immediately before resuming Sudoku.**

Before you begin the game, a practice session of Sudoku will appear. You will have at most 10 minutes to practice. Since this is a practice exercise, error messages will be displayed for feedback purposes as you enter numbers into the squares. PLEASE NOTE: IN THE ACTUAL EXPERIMENT NO ERROR MESSAGES WILL APPEAR.

Scores for the tasks are based on the number of responses completed, so don't worry if you do not have time to finish the tasks.

The system will keep track of time. You will have 18 minutes for Sudoku, 1 and a half minutes for the word task, 48 seconds for each of the two Visual Tasks, and 48 seconds for each of the two Number Series Tasks.

The smaller tasks might be presented in a different order from the ones listed here:

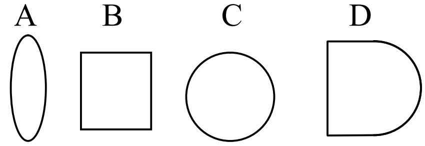
Sudoku Rules: The goal of a Sudoku puzzle is to fill in all the boxes in a 9 X 9 grid, so that each column, each row, and each of the nine 3 X 3 boxes have the numbers 1 – 9 without any of those numbers being repeated.

- Each square must be a number from 1-9.
- All numbers on each row must be unique.
- All numbers on each column must be unique.
- Each 3 X 3 box as well must have a unique number from 1-9.

In the Sudoku, you can temporarily enter more than one number in a square to help you narrow it down. Remember to remove the extra numbers before submitting your answer.

Word Rules: You need to make 20 words (of 3, 4, or 5 letters) using the letters of the word provided.
Example: If the word was 'ARMY', it has the word 'RAM' in it as well as the word 'MAY'.

Visual Rules: "Odd One Out" Choose the shape that doesn't fit the same pattern as the other shapes. Example:



B would be the correct answer, since B is the only shape without any round edges.

Number Series Rules: Guess the missing number in a series of numbers.
Example: 2 4 8 __ The missing number is 16, since each number in the series is doubled.

Game Instructions – Discretionary Multitasking Condition

You will be completing a Sudoku puzzle, along with five smaller tasks. If you finish Sudoku early there is a **SECOND SUDOKU** puzzle below it. Each task is presented on a separate tab. You are free to solve the tasks in any order you like. To change tasks, just click on another tab. The system will keep track of how much time you spend on each task and will time you out when the time limit is reached.
NOTE: You are allowed to switch tabs at any point. Switching to another tab will stop the clock in the task that you left.

Before you begin the game, a practice session of Sudoku will appear. You will have at most 10 minutes to practice. Since this is a practice exercise, error messages will be displayed for feedback purposes as you enter numbers into the squares. PLEASE NOTE: IN THE ACTUAL EXPERIMENT NO ERROR MESSAGES WILL APPEAR.

Scores for the tasks are based on the number of responses completed, so don't worry if you do not have time to finish the tasks.

The system will keep track of time. You will have 18 minutes for Sudoku, 1 and a half minutes for the word task, 48 seconds for each of the two Visual Tasks, and 48 seconds for each of the two Number Series Tasks.

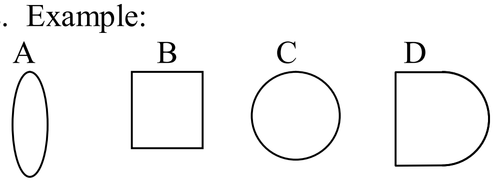
The smaller tasks might be presented in a different order from the ones listed here:
Sudoku Rules: The goal of a Sudoku puzzle is to fill in all the boxes in a 9 X 9 grid, so that each column, each row, and each of the nine 3 X 3 boxes have the numbers 1 – 9 without any of those numbers being repeated.

- Each square must be a number from 1-9.
- All numbers on each row must be unique.
- All numbers on each column must be unique.
- Each 3 X 3 box as well must have a unique number from 1-9.

In the Sudoku, you can temporarily enter more than one number in a square to help you narrow it down. Remember to remove the extra numbers before submitting your answer.

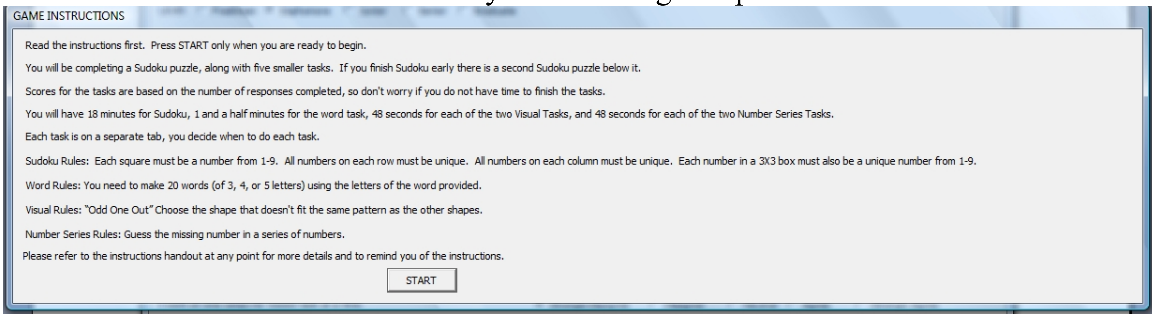
Word Rules: You need to make 20 words (of 3, 4, or 5 letters) using the letters of the word provided. For example: If the word was 'ARMY', it has the word 'RAM' in it as well as the word 'MAY'.

Visual Rules: "Odd One Out" Choose the shape that doesn't fit the same pattern as the other shapes. Example:



B would be the correct answer, since B is the only shape without any round edges.
Number Series Rules: Guess the missing number in a series of numbers.
Example: 2 4 8 __ The missing number is 16, since each number in the series is doubled.

On Screen Instructions – Discretionary Multitasking Sample


GAME INSTRUCTIONS

Read the instructions first. Press START only when you are ready to begin.

You will be completing a Sudoku puzzle, along with five smaller tasks. If you finish Sudoku early there is a second Sudoku puzzle below it.

Scores for the tasks are based on the number of responses completed, so don't worry if you do not have time to finish the tasks.

You will have 18 minutes for Sudoku, 1 and a half minutes for the word task, 48 seconds for each of the two Visual Tasks, and 48 seconds for each of the two Number Series Tasks.

Each task is on a separate tab, you decide when to do each task.

Sudoku Rules: Each square must be a number from 1-9. All numbers on each row must be unique. All numbers on each column must be unique. Each number in a 3X3 box must also be a unique number from 1-9.

Word Rules: You need to make 20 words (of 3, 4, or 5 letters) using the letters of the word provided.

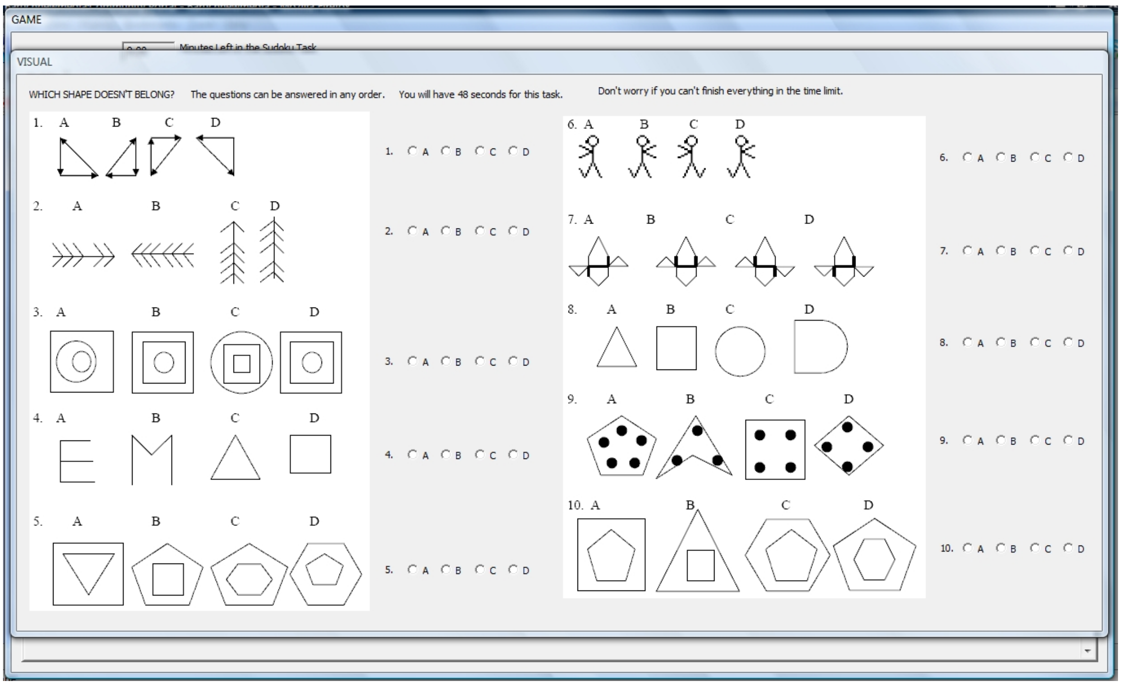
Visual Rules: "Odd One Out" Choose the shape that doesn't fit the same pattern as the other shapes.

Number Series Rules: Guess the missing number in a series of numbers.

Please refer to the instructions handout at any point for more details and to remind you of the instructions.

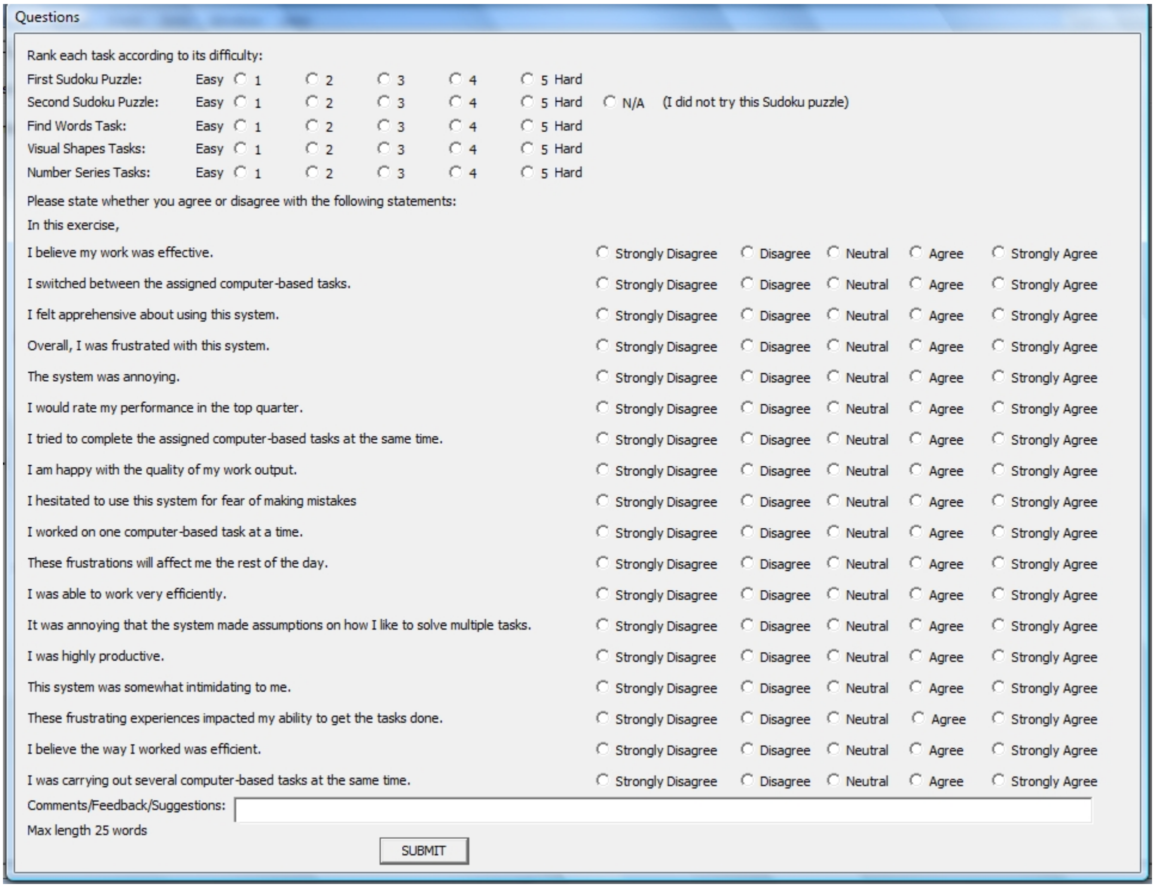
START

Updated Forced Multitasking Environment: larger window size for smaller task.



Appendix I

New Post-test Questionnaire:



Questions

Rank each task according to its difficulty:

First Sudoku Puzzle:	Easy	1	2	3	4	5 Hard	
Second Sudoku Puzzle:	Easy	1	2	3	4	5 Hard	N/A (I did not try this Sudoku puzzle)
Find Words Task:	Easy	1	2	3	4	5 Hard	
Visual Shapes Tasks:	Easy	1	2	3	4	5 Hard	
Number Series Tasks:	Easy	1	2	3	4	5 Hard	

Please state whether you agree or disagree with the following statements:

In this exercise,

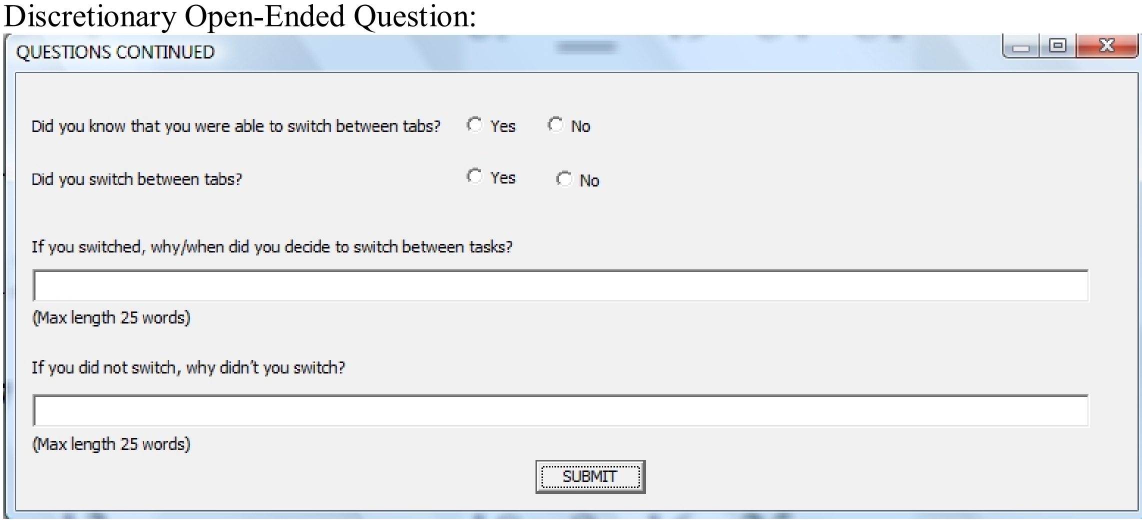
I believe my work was effective.	Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree
I switched between the assigned computer-based tasks.	Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree
I felt apprehensive about using this system.	Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree
Overall, I was frustrated with this system.	Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree
The system was annoying.	Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree
I would rate my performance in the top quarter.	Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree
I tried to complete the assigned computer-based tasks at the same time.	Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree
I am happy with the quality of my work output.	Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree
I hesitated to use this system for fear of making mistakes	Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree
I worked on one computer-based task at a time.	Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree
These frustrations will affect me the rest of the day.	Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree
I was able to work very efficiently.	Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree
It was annoying that the system made assumptions on how I like to solve multiple tasks.	Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree
I was highly productive.	Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree
This system was somewhat intimidating to me.	Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree
These frustrating experiences impacted my ability to get the tasks done.	Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree
I believe the way I worked was efficient.	Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree
I was carrying out several computer-based tasks at the same time.	Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree

Comments/Feedback/Suggestions:

Max length 25 words

SUBMIT

Discretionary Open-Ended Question:



QUESTIONS CONTINUED

Did you know that you were able to switch between tabs? Yes No

Did you switch between tabs? Yes No

If you switched, why/when did you decide to switch between tasks?

(Max length 25 words)

If you did not switch, why didn't you switch?

(Max length 25 words)

SUBMIT

Appendix J

Appendix A. NEW CONSENT FORM 2/6/11

Project Title: Performance Effects of Computer-Based Multitasking Behavior

Principal Investigators:
Raquel Benbunan-Fich (Baruch College) and Rachel Adler (Department of Computer Science, CUNY Graduate Center)

Informed Consent Form

The primary purpose of this study is to examine the performance of users on multiple unrelated computer-based tasks. Participants will be asked to complete three tasks. By participating in this study, you will help advance research and have fun playing games. We anticipate many sessions that will take place and require about an hour per session. A subject can only participate in one session and must be at least 18 years old to participate in this study.

Subjects do not need to bring or have anything to participate in the study. All instructions will be distributed and fully explained before the actual experiment. Participation is completely voluntary. You may discontinue participation at any time. Refusal to participate will involve no penalty. **Participation in this experiment fulfills 1 hr. of your CIS2200 Research Requirement.**

The risks from participating in this study are no more than the risk when playing any computer-based game. The benefit of your participation is enjoyment. In addition, you will help enhance research in human-computer interaction.

All records of this study will be kept confidential. No one other than the PI and Co-PI (Raquel Benbunan-Fich and Rachel Adler) will have access to the data, which will be archived in a secure location after use. If desired, you may see the data collected from only your participation. Any resulting publications from this study will not identify individual participants but will refer to aggregate results.

If you have any questions regarding this research, you can call Raquel Benbunan-Fich at 646-312-3375 or rbfich@baruch.cuny.edu or Rachel Adler at 347-528-4067 or rachelfadler@gmail.com. If you have any questions concerning your rights as a participant in this study, you can contact Keisha Peterson, IRB Administrator, Baruch College/City University of New York at (646)-312-3785.

By signing below, you understand and accept the terms of this research study as stated above, that you are at least 18 years old, and that your participation is completely voluntary.

____________________________ ________________________

Student Signature Investigator's Signature

____________________________ ________________________

Full name (please print) Date

Page 3 of 2

References

ACM SIGCHI Curricula for Human-Computer Interaction. 1992. Retrieved March 8, 2009 from http://sigchi.org/cdg/cdg2.html#2_1.

Adamczyk, P. D. and B. P. Bailey. "If not now, when?: The effects of interruption at different moments within task execution." In *Proceedings of the SIGCHI conference on Human factors in computing systems*, 2004, pp. 271-278.

Bailey, B. P. and S. T. Iqbal. "Understanding changes in mental workload during execution of goal-directed tasks and its application for interruption management." *ACM Transactions on Computer-Human Interaction* (14:4), 2008, pp. 1-28.

Bailey, B. P. and J. A. Konstan. "On the need for attention-aware systems: Measuring effects of interruption on task performance, error rate and affective state." *Journal of Computers in Human Behavior* (22:4), 2006, pp. 658-708.

Bannon, L. J. "Issues in design: Some notes," in *User Centered System Design*, Norman D. A. and Draper S. W. (Eds.) Lawrence Erlbaum, Hillsdale, N.J., USA, 1986, pp. 25-29.

Bardram, J., J. Bunde-Pedersen and M. Soegaard. "Support for activity-based computing in a personal computing operating system." In *Proceedings of the SIGCHI conference on Human Factors in computing systems, CHI '06*, 2006, pp. 211-220.

Bedny, G. and W. Karkowsky. *A Systemic-Structural Theory of Activity: Applications to Human Performance and Work Design*, Boca Raton, FL: CRC/Taylor & Francis, 2007.

Bell, C. S., D. R. Compeau and F. Olivera. "Understanding the social implications of technological multitasking: A conceptual model." *HCI Research in MIS*, 2005, pp. 80-84.

Bellotti, V., N. Ducheneaut, M. Howard, I. Smith and R. E. Grinter. "Quality versus quantity: e-mail-centric task management and its relation with overload." *Human-Computer Interaction* (20:1/2), 2005, pp. 89-138.

Benbunan-Fich, R., R. F. Adler and T. Mavlanova. "Patterns of home computer usage among college students: An exploratory study." In *Proceedings of the AMCIS 2009*, 2009a, pp. 1-9, #112.

Benbunan-Fich, R., R. F. Adler and T. Mavlanova. "Towards new metrics for multitasking behavior." In *Proceedings of the CHI '09 extended abstracts on human factors in computing systems*, 2009b, pp. 4039-4044.

Benbunan-Fich, R., R. F. Adler and T. Mavlanova. "Measuring multitasking behavior with activity-based metrics." *ACM Transactions on Computer-Human Interaction* (18:2), 2011, pp. 1-22.

Benbunan-Fich, R. and G. E. Truman. "Multitasking with laptops during meetings." *Communications of the ACM* (52:2), 2009, pp. 139-141.

Bluedorn, A. C., C. F. Kaufman and P. M. Lane. "How many things do you like to do at once? An introduction to monochronic and polychronic time." *Academy of Management Executive* (6:4), 1992, pp. 17-26.

Card, S., T. P. Moran and A. Newell. *The psychology of human computer interaction*, Hillsdale, N.J., USA: Lawrence Erlbaum Associates, Inc., 1983.

Card, S. K. and A. Henderson. "A multiple, virtual-workspace interface to support user task switching." In *Proceedings of the SIGCHI/GI conference on human factors in computing systems and graphics interface*, 1987, pp. 53-59.

Ceaparu, I., J. Lazar, K. Bessiere, J. Robinson and B. Shneiderman. "Determining causes and severity of end-user frustration." *International Journal of Human-Computer Interaction* (17:3), 2004, pp. 333-356.

Crook, C. and D. Barrowcliff. "Ubiquitous computing on campus: Patterns of engagement by university students." *International Journal of Human-Computer Interaction* (13:2), 2001, pp. 245 - 256.

Cutrell, E. B., M. Czerwinski and E. Horvitz. "Effects of instant messaging interruptions on computing tasks." In *Proceedings of the CHI '00 extended abstracts on Human factors in computing systems*, 2000, pp. 99-100.

Cypher, A. "The structure of users' activities," in *User centered system design*, D. A. Norman and S. W. Draper (Eds.) Lawrence Erlbaum, Hillsdale, N.J., USA, 1986, pp. 243-263.

Czerwinski, M., E. Horvitz and S. Wilhite. "A diary study of task switching and interruptions." In *Proceedings of the SIGCHI conference on human factors in computing systems*, 2004, pp. 175-182.

Czerwinski, M., G. Robertson, B. Meyers, G. Smith, D. Robbins and D. Tan. "Large display research overview." In *Proceedings of the CHI '06 extended abstracts on human factors in computing systems*, 2006, pp. 69-74.

Dix, A., J. Finlay, G. D. Abowd and R. Beale. *Human-computer interaction (Third Edition)*, Prentice Hall, 2004.

Douglas, S., M. Tremaine, L. Leventhal, C. E. Wills and B. Manaris. "Incorporating human-computer interaction into the undergraduate computer science curriculum." In *Proceedings of the 33rd SIGCSE technical symposium on Computer science education*, 2002, pp. 211-212.

Draper, S. W. and D. A. Norman. "Introduction," in *User centered system design*, D. A. Norman and S. W. Draper (Eds.), Hillsdale, N.J., USA, 1986, pp. 1-6.

Gonzalez, V. M. and G. Mark. "'Constant, constant, multi-tasking craziness': Managing multiple working spheres." In *Proceedings of the SIGCHI conference on Human factors in computing systems*, 2004, pp. 113-120.

Goodwin, N. C. "Functionality and usability." *Communications of the ACM* (30:3), 1987, pp. 229-233.

Hair, J. F., R. E. Anderson, R. L. Tatham and W. C. Black. *Multivariate Analysis*, Upper Saddle River, NJ: Prentice Hall, 1998.

Hembrooke, H. and G. Gay. "The laptop and the lecture: The effects of multitasking in learning environments." *Journal of Computing in Higher Education* (15:1), 2003, pp. 46-64.

Hochheiser, H. and J. Lazar. "HCI and societal Issues: A framework for engagement." *International Journal of Human-Computer Interaction* (23:3), 2007, pp. 339 - 374.

Iqbal, S. T. and B. P. Bailey. "Leveraging characteristics of task structure to predict the cost of interruption." In *Proceedings of the PSIGCHI conference on Human Factors in computing systems, CHI '06*, 2006, pp. 741-750.

Iqbal, S. T. and B. P. Bailey. "Understanding and developing models for detecting and differentiating breakpoints during interactive tasks." In *Proceedings of the SIGCHI conference on Human factors in computing systems, CHI '07*, 2007, pp. 697-706.

Iqbal, S. T. and B. P. Bailey. "Effects of intelligent notification management on users and their tasks." In *Proceedings of the twenty-sixth annual SIGCHI conference on human factors in computing systems, CHI '08*, 2008, pp. 93-102.

Iqbal, S. T. and E. Horvitz. "Conversations amidst computing: A study of interruptions and recovery of task activity." In *Proceedings of the 11th international conference on User Modeling*, 2007a, pp. 350-354.

Iqbal, S. T. and E. Horvitz. "Disruption and recovery of computing tasks: field study, analysis, and directions." In *Proceedings of the SIGCHI conference on Human factors in computing systems*, 2007b, pp. 677-686.

Isaacs, E., C. Kamm, D. J. Schiano, A. Walendowski and S. Whittaker. "Characterizing instant messaging from recorded logs." In *Proceedings of the CHI '02 extended abstracts on Human factors in computing systems*, 2002, pp. 720-721.

Jin, J. and L. A. Dabbish. "Self-interruption on the computer: a typology of discretionary task interleaving." In *Proceedings of the 27th international conference on human factors in computing systems, CHI '09*, 2009, pp. 1799-1808.

John, B. E. and J. H. Morris. "HCI in the school of computer science at Carnegie Mellon University." In *Proceedings of the INTERACT '93 and CHI '93 conference on Human factors in computing systems*, 1993, pp. 49-50.

Johnson, P., J. May and H. Johnson. "Introduction to multiple and collaborative tasks." *ACM Transactions on Computer-Human Interaction* (10:4), 2003, pp. 277-280.

Karam, M. and M. C. Schraefel. "A study on the use of semaphoric gestures to support secondary task interactions." In *Proceedings of the CHI '05 extended abstracts on Human factors in computing systems*, 2005, pp. 1961-1964.

Kleinman, L. "Perceived productivity and the social rules for laptop use in work meetings." In *Proceedings of the CHI '09 extended abstracts on Human factors in computing systems*, 2009, pp. 3895-3900.

Lee, H. "Time and information technology: monochronicity, polychronicity and temporal symmetry." *European Journal of Information Systems* (8:1), 1999, pp. 16-26.

Mark, G., V. M. Gonzalez and J. Harris. "No task left behind?: Examining the nature of fragmented work." In *Proceedings of the SIGCHI conference on Human factors in computing systems*, 2005, pp. 321-330.

Mark, G., D. Gudith and U. Klocke. "The cost of interrupted work: more speed and stress." In *Proceedings of the twenty-sixth annual SIGCHI conference on human factors in computing systems, CHI '08*, 2008, pp. 107-110.

Matthews, T., M. Czerwinski, G. Robertson and D. Tan. "Clipping lists and change borders: improving multitasking efficiency with peripheral information design." In *Proceedings of the SIGCHI conference on Human Factors in computing systems*, 2006, pp. 989-998.

McCrickard, D. S., R. Catrambone, C. M. Chewar and J. T. Stasko. "Establishing tradeoffs that leverage attention for utility: empirically evaluating information display in notification systems." *International Journal of Human-Computer Studies* (58:5), 2003a, pp. 547-582.

McCrickard, D. S., C. M. Chewar, J. P. Somervell and A. Ndiwalana. "A model for notification systems evaluation-assessing user goals for multitasking activity." *ACM Transactions on Computer-Human Interaction* (10:4), 2003b, pp. 312-338.

McCrickard, D. S., M. Czerwinski and L. Bartram. "Introduction: design and evaluation of notification user interfaces." *International Journal of Human-Computer Studies* (58:5), 2003c, pp. 509-514.

McFarlane, D. "Comparison of four primary methods for coordinating the interruption of people in human-computer interaction." *Human-Computer Interaction* (17:1), 2002, pp. 63-139.

McFarlane, D. C. and K. A. Latorella. "The scope and importance of human interruption in human-computer interaction design." *Human-Computer Interaction* (17:1), 2002, pp. 1-61.

Miyata, Y. and D. A. Norman. "Psychological issues in support of multiple activities," in *User Centered System Design*, D. A. Norman and S. W. Draper (Eds.) Lawrence Erlbaum, Hillsdale, N.J., USA, 1986, pp. 265-284.

Nagata, S. F. "Multitasking and interruptions during mobile web tasks." In *Proceedings of the 47th Annual Meeting of the Human Factors and Ergonomics Society*, 2003, pp. 1341-1345.

Norman, D. A. *The design of everyday things*, Basic Books, 1988.

Norman, D. A. and S. W. Draper. *User centered system design*, Hillsdale, N.J., USA: Lawrence Erlbaum, 1986.

O'Conaill, B. and D. Frohlich. "Timespace in the workplace: dealing with interruptions." In *Proceedings of the Conference Companion on Human Factors in Computing Systems, CHI '95*, 1995, pp. 262-263.

Ophir, E., C. Nass and A. D. Wagner. "Cognitive control in media multitaskers." In *Proceedings of the National Academy of Sciences*, 2009, pp. 1-5.

Oulasvirta, A. and P. Saariluoma. "Long-term working memory and interrupting messages in human-computer interaction." *Behaviour and Information Technology* (23:1), 2004, pp. 53-64.

Oulasvirta, A. and P. Saariluoma. "Surviving task interruptions: Investigating the implications of long-term working memory theory." *International Journal of Human-Computer Studies* (64:10), 2006, pp. 941-961.

Palladino, L. J. *Find Your Focus Zone: An Effective New Plan to Defeat Distraction and Overload*, New York: Free Press, 2007.

Payne, S. J., G. B. Duggan and H. Neth. "Discretionary task interleaving: Heuristics for time allocation in cognitive foraging." *Journal of Experimental Psychology: General* (136:3), 2007, pp. 370-388.

Reichman, R. "Communication paradigms for a window system," in *User centered system design*, D. A. Norman and S. W. Draper (Eds.) Lawrence Erlbaum, Hillsdale, N.J., USA, 1986, pp. 285-313.

Renaud, K., J. Ramsay and M. Hair. "'You've got e-mail!'... Shall I deal with it now? Electronic mail from the recipient's perspective." *International Journal of Human-Computer Interaction* (21:3), 2006, pp. 313-332.

Salvucci, D. D., N. A. Taatgen and J. P. Borst. "Toward a unified theory of the multitasking continuum: from concurrent performance to task switching, interruption, and resumption." In *Proceedings of the CHI '09 Human factors in computing systems*, 2009, pp. 1819-1828.

Sanderman, A., J. Sturm, E. Den Os, L. Boves and A. Cremers. "Evaluation of the Dutch Train Timetable Information System developed in the Arise project." In *Proceedings of the 4th IEEE workshop on interactive voice technology for telecommunications applications*, 1998, pp. 91-96.

Scupelli, P., S. Kiesler, S. R. Fussell and C. Chen. "Project view IM: A tool for juggling multiple projects and teams." In *Proceedings of the CHI '05 extended abstracts on Human factors in computing systems*, 2005, pp. 1773-1776.

Shneiderman, B. *Designing the user interface. Strategies for effective human-computer interaction (1st edition)*, Addison-Wesley, 1992.

Silver, M. S. "Decision Support Systems: Directed and Nondirected Change." *Information Systems Research* (1:1), 1990, pp. 47-70.

Speier, C., I. Vessey and J. S. Valacich. "The effects of interruptions, task complexity, and information presentation on computer-supported decision-making performance." *Decision Sciences* (34:4), 2003, pp. 771-797.

Staples, D. S., J. S. Hulland and C. A. Higgins. "A self-efficacy theory explanation for the management of remote workers in virtual organizations." *Organization Science* (10:6), 1999, pp. 758-776.

Su, N. M. and G. Mark. "Communication chains and multitasking." In *Proceedings of the SIGCHI conference on Human factors in computing systems*, 2008, pp. 83-92.

Thatcher, J. B. and P. L. Perrewe. "An empirical examination of individual traits as antecedents to computer anxiety and computer self-efficacy." *MIS Quarterly* (26:4), 2002, pp. 381-396.

Trafton, J. G., E. M. Altmann, D. P. Brock and F. E. Mintz. "Preparing to resume an interrupted task: Effects of prospective goal encoding and retrospective rehearsal." *International Journal of Human-Computer Studies* (58:5), 2003, pp. 583-603.

Voida, S. and E. D. Mynatt. "It feels better than filing: everyday work experiences in an activity-based computing system." In *Proceedings of the 27th international conference on Human factors in computing systems, CHI '09*, 2009, pp. 259-268.

Wasson, C. "Multitasking during virtual meetings." *Human Resource Planning* (27:4), 2004, pp. 47-61.

Weisz, J. D., S. Kiesler, H. Zhang, Y. Ren, R. E. Kraut and J. A. Konstan. "Watching together: integrating text chat with video." In *Proceedings of the SIGCHI conference on Human factors in computing systems*, 2007, pp. 877-886.

Wild, P. J., P. Johnson and H. Johnson. "Towards a composite modelling approach for multitasking." In *Proceedings of the 3rd annual conference on task models and diagrams*, 2004, pp. 17-24.

Zhang, P., F. Nah and J. Preece. "HCI studies in management information systems." *Behaviour and Information Technology* (23:3), 2004, pp. 147-151.

Zhang, Y., R. S. Goonetilleke, T. Plocher and S.-F. M. Liang. "Time-related behaviour in multitasking situations." *International Journal of Human-Computer Studies* (62:4), 2005, pp. 425-455.